SOME PRACTICAL ISSUES IN THE DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS

Tony Sahama

THIS THESIS IS PRESENTED IN FULFILMENT OF
THE REQUIREMENTS OF THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF COMPUTER SCIENCE AND MATHEMATICS
FACULTY OF SCIENCE, ENGINEERING AND TECHNOLOGY

VICTORIA UNIVERSITY OF TECHNOLOGY

2003

Certificate of Originality

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgement is made in the text.

I also hereby declare that this thesis is written in accordance with the University's Policy with respect to the Use of Project Reports and Higher Degree Theses.

T. Sahama

Abstract

Deterministic computer simulations of physical experiments are now common techniques in science and engineering. Often, physical experiments are too time consuming, expensive or impossible to conduct. Complex computer models or codes, rather than physical experiments lead to the study of computer experiments, which are used to investigate many scientific phenomena of this nature. A computer experiment consists of a number of runs of the computer code with different input choices. The Design and Analysis of Computer Experiments is a rapidly growing technique in statistical experimental design.

This thesis investigates some practical issues in the Design and Analysis of Computer experiments and attempts to answer some of the questions faced by experimenters using computer experiments. In particular, the question of the number of computer experiments and how they should be augmented is studied and attention is given to when the response is a function over time.

Investigation of the appropriate sample size for computer experiments is undertaken using three fire models and one circuit simulation model for empirical validation. Detailed illustrations and some guidelines are given for the sample size and an empirical relationship is established showing how the average prediction error from a computer experiment is related to the sample size.

When the average prediction error following a computer experiment is too large the question of how to augment the computer experiment is raised. Two approaches are studied, evaluated and compared. The first approach involves adding one point at a time and choosing that point with the maximum predicted variance, while the second approach involves maximising the determinant of the variance-covariance matrix of the prediction errors of a candidate set.

Rather than just examining a whole series of practical cases, the machinery of computer experiments is also used to study computer experiments themselves. The inputs of the model are the parameters of the *Kriging* model as well as the number of input runs while the output is a measure of the prediction error. This study provides predictions of the average prediction error for a wide range of computer models.

Many computer codes provide not just a univariate response but a trace of responses at various values of a time parameter. A method for analysing such computer experiments is proposed and illustrated.

Acknowledgements

- I wish to express my sincere gratitude to my supervisor, Dr. Neil Diamond, for his guidance, constructive criticism, valuable genuine advice and encouragement given throughout the period of my Ph.D. study at Victoria University, Melbourne.

- I also wish to extend my deepest appreciation to Assoc. Prof. Neil Barnet, former Head of the School, for providing me with a departmental scholarship for the successful completion of this study.

- Thanks are also rendered to Mr. P. Rajendran, Damon Burgess, Rowan Macintosh and other colleagues who have been of immense assistance throughout this study.

- I also would like to record my grateful thanks to Mr. David Abercrombie and Mr. Ken Ling, School of Information Systems, Faculty of Information Technology, Queensland University of Technology for their encouragement and support in numerous ways.

- Special thanks to my wife Charlotte and our children Ishani and Ishara for giving me all the support throughout my studies.

Dedication

To my parents who taught me to work hard, to Charlotte who guided me to be fair and stay curious, and to Ishani and Ishara for keeping me honest.

Preface

Parts of Chapters 3 and 4 appeared in Sahama and Diamond (2001).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Computer Experiments

The advancement of high-speed computers has made experimentation via computer modelling common in many areas of science and technology. Computer modelling is having a significant impact on scientific research. Virtually every area of science and technology is affected. A computer model or simulator usually involves complicated or high dimensional mathematical functions. Based on the mathematical formulation, the computer model or code produces outputs, if the required values of the input variables are provided. Running the computer simulation can be expensive in different ways. It can be labour intensive and/or time consuming. If the computer simulator is expensive to run, a natural strategy is to build a predictor from relatively few runs to act as a computationally less expensive surrogate (Welch *et al.* 1992) which can be used in a variety of ways, for example during optimisation of the output.

In contrast, many complex processes that are conducted as physical experimentation are too time consuming and expensive (Sacks *et al.* 1989b). Moreover, for many systems such as global weather modeling, environmental modelling and fire modelling, physical experimentation may simply be impossible. As a result, experimenters have increasingly moved to use mathematical models to simulate these complex systems. Enhancement of computer power has permitted both

greater complexity and more extensive use of such models in scientific experimentation as well as in industrial processes. Computer simulation is invariably cheaper than physical experimentation although these codes can be computationally demanding (Welch and Sacks, 1991).

This computer experiment approach has opened up a new avenue, "Design and Analysis of Computer Experiments" (DACE), which is somewhat different to the traditional methodology for the "Design of Experiments" (DoE). A significant comparison between DACE and DoE concepts was presented by Booker *et al.* (1996). Booker (1996) compared DACE to DoE in three application areas: electrical power system design, aeroelastic design and aerodynamic design. As Booker suggested *"DoE and DACE ... [are basically] the same approach employing different tools as appropriate"*.

In general, computer models or codes consist of multivariate inputs, which can be scalars or functions (Sacks *et al.* 1989b) and the resulting output from the same code may also be univariate and/or multivariate. In addition, the output can be a time dependent function and the enhanced ability to gather and analyse a number of summary responses was highlighted by Sacks *et al.* (1989b). The input dimension differs according to the purpose and basis of the original computer model. Selecting a number of runs out of various input configurations results in *Computer Experiments*.

One of the important application areas is the computer simulation of integrated circuits. Here $(\mathbf{x})$ defines various circuit parameters, such as transistor characteristics, and (y) is a measurement of the circuit performance, such as output voltage. The literature shows some other applications in a wide variety of fields such as plant ecology (Bartell *et al.*, 1981 and 1983), Heat combustion (Miller and Frenklach, 1983), chemometrics (Ho *et al.*, 1984), controlled-nuclear-fusion devices (Nassif *et al.*, 1984), thermal-energy storage (Currin *et al.* 1988), VLSI-circuit design (Sharifzadeh *et al.*, 1989), Solar Collector Experiments and Automotive Industry (Schonlau, 1997), Biomechanical Engineering (Chang *et al.* 1999) and Oil hydrocarbon reservoir simulation (Graig *et al.*, 1996) and (Kennedy and O'Hagan, 2000). Other application areas, as highlighted in Koehler (1990),

were a mold filling process for manufacturing automobiles, chemical kinetic models, a thermal energy storage model and the transport of polycyclic aromatic hydrocarbon spills in streams using structured activity relationships model being of use in plant ecology. In fact, the widespread use of computer models and experiments for simulating real phenomena generates examples in virtually all areas of science and engineering.

1.2 Role of Experimental Designs in DACE

Experimental design, as a statistical discipline, began with the pioneering work of R.A. Fisher in the 1920s. It is one of the most powerful tools of statistics and is widely used when designing experiments. Experimental designs have been used to mitigate the effect of random noise in experimental outcomes. However, the key ideas of randomisation, replication, and blocking are not useful for computer experiments.

In contrast, the Design of Computer Experiments is a new avenue in the Design of Experiments since there are no random errors. Also, replication is not necessary since one always obtains the same response at the same input settings (Booker 1996). The role of statistical experimental design in Computer experiments was reviewed by Sacks *et al.* (1989b), stating that .... *"the selection of inputs at which to run a computer code is still an experimental design problem"*.

1.3 Differences between Computer Experiments and Other Experimental Design

Although Factorial and Fractional Factorial designs are most commonly used for physical experiments, in computer experiments perhaps the most common designs are Latin Hypercube Designs (LHD). These were the first type of designs to be explicitly considered as experimental designs for deterministic computer codes.

Computer models are deterministic: replicate observations from running the

code with the same inputs will be identical. Since these models have no random or measurement errors, computer experiments are different from physical experiments calling for distinct techniques for design. These deterministic computer experiments differ substantially from the majority of physical experiments traditionally performed by scientists. Such experiments usually have substantial random error due to variability in the experimental units (Sacks *et al.* 1989b).

The remarkable methodology of statistical design of experiments was introduced in the 1920's and popularised among scientists following the publication of Fisher (1935). The associated analysis of variance is a systematic way of separating important treatment effects from the background noise (as well as from each other). Fisher's methods of blocking, replication and randomization in these experiments reduced the effect of random error, provided valid estimates of uncertainty and preserved the simplicity of the models. The deterministic computer codes considered in this thesis differ from codes in the simulation literature (Sacks *et al.* 1989b), which incorporate substantial random error through random number generators. The random (as opposed to deterministic) simulation, i.e., computer models that use pseudorandom numbers (as in many telecommunications and logistic applications) described by Law and Kelton (2000) is a significant contribution for simulation experiments. It has been natural, therefore, to design and analyse such stochastic simulation experiments using standard techniques for physical experiments. However, it is doubtful whether these methods are appropriate for computer experiments considered here since the prediction will not then match the observed deterministic response. For this reason other methods of design and analyses have been developed by a number of authors.

The design problem is the choice of inputs for efficient analysis of the data. A computer experimental design consists of a set of sample points to be evaluated through the computer code or model. The observations from this design are then used to develop a computationally cheaper surrogate model to the code. This new model is used to estimate the code inexpensively and efficiently, to investigate the behavior of the function, or to optimise some aspect of the function. "*Computer experiments are efficient methods of extracting information about the*

unknown function and providing an approximating model that can be inexpensively evaluated" (Koehler and Owen, 1996).

1.4 Designs for Computer Experiments

McKay *et al.* (1979), were the first to explicitly consider experimental design for deterministic computer codes. With the input variable given by $\mathbf{X} = (X^1, \ldots, X^d)$ where X^j is a standardised input between 0 and 1, and the output produced by the computer code given by $y = h(\mathbf{X})$, they compared 3 methods of selecting the input variables:

1. Random Sampling

2. Stratified Sampling

3. Latin Hypercube Sampling - an extension of stratified sampling which ensures that each of the input variables has all portions of its range represented. A uniform Latin Hypercube sample of size n has

$$\tilde{X}_i^j = \frac{\pi_j(i) - U_i^j}{n}, 1 \leq i \leq n, 1 \leq j \leq d$$

where $\pi_j(1), \ldots, \pi_j(n)$ are random permutations of the integers $1, \ldots, n$, $U_i^j \sim U[0,1]$ and the d permutations and nd uniform variates are mutually independent (Owen, 1992a). Many authors use the simpler Lattice sample, following Patterson (1954), where

$$\tilde{X}_i^j = \frac{\pi_j(i) - \frac{1}{2}}{n}, \ 1 \leq i \leq n, 1 \leq j \leq d.$$

A slightly altered definition has been used in this thesis.

Latin Hypercube samples have been used extensively. Iman and Conover (1980) applied Latin Hypercube Sampling to a Ground Water Flow System where it was desired to consider the potential escape of radio nucleotides from a deposit for radioactive waste and their migration from the subsurface to the surface environment. They also gave a generalisation of Latin Hypercube Sampling that

allowed for different assumptions on the input variable to be studied without additional computer runs. Iman and Conover (1982) showed how Latin Hypercube Samples could be modified to incorporate the correlations that may exist among the input variables. The method was used on a model for the study of Geological Disposal of Radioactive Waste. Stein (1987) gave the asymptotic variance of the Latin Hypercube based estimation and showed that the estimate of the expected value of a function based on a Latin Hypercube sample is asymptotically normal and that the improvement over simple random sampling depends on the degree of additivity of the function of the input variables. He also provided a better method than Iman and Conover (1982), of producing Latin Hypercube sampling when the input variables are dependent.

Other experimental designs based on different optimality criteria were studied by Koehler (1990) who considered Entropy, Mean Square Error, Minimax, Maximin and Star-discrepancy based designs. In addition, three examples were studied: one a chemical kinetics problem involving 11 differential equations. The second example investigated was a large linear system of differential equations involving methane combustion with seven rate constants regarded as inputs. To reduce the computer time, the design is restricted to a central composite design type (Box and Draper, 1987) but with unknown shrinkage factors for the cube points and the star points of the design.

A number of computational methods for augmenting the design were described which would be classified as single stage methods, sequential methods without adaptation to the data, and sequential methods with adaptation. The problem of how to design an augmenting experiment is an important one. It is clear, however, that much more work needs to be done to make these methods easier to use and this is something this research study aims to accomplish.

Owen (1992b) generalised Latin Hypercube samples by using orthogonal arrays (Raghavarao, 1971). An orthogonal array of strength t is a matrix of n rows and k columns with elements taken from a set of q symbols such that in any n by t matrix each of the q^t possible rows occurs the same number λ of times. Such an array is denoted by $\mathrm{OA}(n, k, q, t)$. A generalisation of the Latin Hypercube

sample is given by

$$X_i^j = \frac{\pi_j(A_i^j) + U_i^j}{q}, \ 0 \leq i \leq 1, 0 \leq j \leq -1$$

where the π_j are independent permutations of $0, \ldots, q-1$ and A_i^j is the value of the i^{th} row and j^{th} column of an orthogonal array. Similarly, Lattice samples can be generalised by taking

$$X_i^j = \frac{\pi_j(A_i^j) + \frac{1}{2}}{q}.$$

Owen (1992b) suggests that the arrays OA$(q^2, k, q, 2)$ are a good choice for computer experiments since the $n = q^2$ points plot as a $q \times q$ grid on each bivariate margin. He also presents OA of strength 2 of the form OA $(q^2, q+1, q, 2)$ for $q = 2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, 19, 25, 27,$ and 32 and OA $(2q^2, 2q+1, q, 2)$ for $q = 2, 3, 4, 5, 7, 8, 9, 11, 13, 16$ although the latter designs include the same projections in three columns that include repeat runs, an undesirable feature for a computer experiment. In a later paper, (Owen 1994b), it is conjectured that sub-arrays of the form OA $(2q^2, 2q, q, 2)$ do not have this defect.

Owen (1992a) also suggests ways to augment a computer experiment. If an experiment based on an OA $(q, d, q, 1)$ has been run then it can be used as an augmenting set for those runs that would complete an OA $(2q, d, q, 1)$. If we want to increase the number of variables then OA $(q^2, d, q, 2) \times$ OA $(q^2, d^2, q^2, 1)$ could be used.

Independent to the work of Owen, Tang (1993) also generalised Latin Hypercubes by developing Orthogonal Array based Latin Hypercube Designs. OA based Latin Hypercubes offers a substantial improvement over Latin Hypercube sampling and are proposed to be a more appropriate design for Computer Experiments.

Booker (1996) compared two experimental designs namely Central Composite Designs for DoE and Space-Filling Designs for DACE applications. According to Booker, Latin Hypercube samples are one class of quasi-Monte Carlo integration designs based on random assignments that are generally ineffective for large problems. Booker suggested that the possibility of using Orthogonal Arrays (OA) as

an experimental design to be used in Computer Experiments for DACE would give more confidence because of its good space-filling properties and infiltration of the design space. He gave examples where he had used OA-based Latin Hypercube samples which are derived from OAs. for an areoelastic simulation of the performance of a helicopter rotor.

1.5 Analysis of Computer Experiments

In a computer experiment, observations are made on a response function y by running a (typically complex) computer model at various choices of input factors $\mathbf{x}$. For example, in the chemical kinetics of methane combustion, $\mathbf{x}$ can be a set of rate constants in a system of differential equations and y can be a concentration of a chemical species some time after combustion. Solving the differential equations numerically for specified $\mathbf{x}$ yields a value for y. Because running the equations solver is expensive, the aim is to estimate the relationship between $\mathbf{x}$ and y from a moderate number of runs so that y can be predicted at untried inputs.

Extracting observations from this design can be used to build up a computationally inexpensive surrogate model to the selected simulator or computer code. This surrogate model is used to approximate the computer simulator or model cheaply and efficiently, to investigate the behaviour of the function and/or to optimise some aspect of the function.

A major contribution to the area was made by Sacks *et al.* (1989a). The response was modeled as:

$$\begin{aligned} \text{Response} &= \text{Linear Model} + \text{Departure} \\ y(\mathbf{x}) &= \sum_{j=1}^{k} \beta_j f_j(x) + z(\mathbf{x}). \end{aligned}$$

The approach taken by Sacks *et al.* (1989a) was to model the systematic departure z as a realization of a stochastic process in which the covariance function of z relates to the smoothness of the response. The covariance of the responses

to two d-dimensional inputs $\mathbf{t} = (t_1.....t_d)$ and $\mathbf{u} = (u_1....u_d)$ is given by.

$$\text{Cov}(z(\mathbf{t}).\, z(\mathbf{u})) \;=\; \sigma_z^2 \prod_{j=1}^{d} R_j(t_j.\, u_j) \tag{1.1}$$

where

$$R_j(t_j.\, u_j) \;=\; \exp(-\theta(t_j - u_j)^2). \tag{1.2}$$

Here $\theta > 0$ defines the correlation structures of z and σ_z^2 is a scale factor. The authors discussed the importance of the parameter θ. When θ is large there is small correlation between observations and therefore prediction is harder. On the other hand when θ is small there is large correlation between observations and prediction is much easier. Selection of the correlation function plays a critical role in the prediction process. Koehler (1990) discussed different correlation families and their suitability to the prediction process and the effects of θ on prediction.

Given the model above the authors derived the Best Linear Predictor and the mean square error of prediction. To come up with a design they try to minimize the integrated mean square error (IMSE) of prediction. The IMSE is given by

$$J_\theta(S.\hat{Y}) \;=\; (1/\sigma_z^2) \int E_\theta(\hat{Y}(\mathbf{x}) - Y(\mathbf{x}))^2 dx$$

where $\hat{Y}(\mathbf{x})$ is the Best Linear Predictor. (S) is the design. the design-prediction strategy is $(S.\ \hat{Y})$ and the integration is over the region of interest.

Since there usually is no way of guessing the value of θ prior to the experiment the strategy adopted by the authors is to choose a value of θ. say θ_A. that gives a design-prediction strategy that performs well for a wide range of true (but unknown) θ_T. A number of different values of θ_A are chosen and for each the optimal design S_{θ_A} and Best Linear Predictor $\hat{Y}_{\theta_A}$ are found. Then for various values of θ_T the IMSE $J_{\theta_T}(S_{\theta_A}.\hat{Y}_{\theta_A})$ is calculated. This quantity is a measure of the performance of the strategy $(S_{\theta_A}.\hat{Y}_{\theta_A})$ when θ_T is true.

A measure of relative efficiency is

$$J_{\theta_I}(S_{\theta_T}.\hat{Y}_{\theta_I})/J_{\theta_T}(S_{\theta_A}.\hat{Y}_{\theta_A}).$$

A robust strategy is to choose θ_A so that these relative efficiencies are as constant as possible.

The paper by Sacks *et al.* (1989b) was also very important. The objectives of a Computer Experiment are set out as:

1. Predicting the response at untried inputs

2. Optimizing a function of the response

3. Validation and Verification (Matching the computer code to physical data). Kleijnen (2000) highlighted the importance of the validation and verification in the simulation experiments. The author has claimed the validation and verification has many facets, including philosophical and mathematical-statistical problems. In particular, the author stressed "In practice, even quite simple simulations are not validated through correct statistical techniques" [see Sargent *et al.*, (2000) for more details]

Sacks *et al.* (1989b) concentrated on the first objective. The basic statistical questions were:

- The design problem:
  At which input "sites" $S = (s_1, \ldots, s_n)$ should the data $y(s_1), \ldots, y(s_n)$ be calculated ?

- The analysis problem:
  How should the data be used to meet the objective ?

They claimed that statistics had a role in Computer Experiments since selection of inputs at which to run a computer code is an Experimental Design problem and the quantification of the uncertainty associated with prediction from fitted models is a statistical problem.

Two rationales for modeling the deterministic departure as a realisation of a stochastic process were advanced:

1. The departure from the **Kriging** model (surrogate) may resemble a path of a suitably chosen stochastic process.

Kriging

Kriging is named after the South-African mining engineer D.G. Krige. It is an interpolation method that predicts unknown values of random function or random process (Cressie, 1993). More precisely, a Kriging prediction is a weighted linear combination of all output values already observed. These weights depend on the distances between the location to be predicted and the locations already observed. Kriging assumes that *the closer the input data are, the more positively correlated the prediction errors are*. This assumption is modeled through the correlogram or the related variogram.

Kriging is popular in *deterministic simulation*. Compared with linear regression analysis, Kriging has an important advantage in deterministic simulation: Kriging is an *exact interpolator*; that is, predicted values at observed input values are exactly equal to the observed (simulated) output values [Kleijnen and van Beers (2003)].

2. $y(.)$ may be regarded as a Bayesian *prior* on the true response function, with the β's either specified *a priori* or given.

The Best Linear Predictor Estimate (Welch *et al.* 1992) was used - this is related to the concept of *Kriging* in the Geostatistical literature (Cressie, 1986). Alternatively the posterior mean could be used from a Bayesian viewpoint.

The correlation function used by Sacks *et al.* (1989b) includes a different θ for each input.

$$R(\mathbf{w}, \mathbf{x}) = \prod_{j=1}^{d} \exp\left(-\theta_j |w_j - x_j|^p\right) \tag{1.3}$$

and the authors also foreshadow the use of a different p_j for each input.

If the model is

$$y(\mathbf{x}) = \beta + z(\mathbf{x}) \tag{1.4}$$

then obtaining the maximum likelihood estimators of $\theta_1, \ldots, \theta_d$, p, β and σ^2 reduces to numerically optimizing

$$-\frac{1}{2}(n \ln \hat{\sigma}^2 + \ln \det \mathbf{R}_D) \tag{1.5}$$

where R_D is the matrix of correlation for the design points.

$$\hat{\beta} = (1^T \mathbf{R}_D^{-1} 1)^{-1} 1^T \mathbf{R}_D^{-1}(\mathbf{y}) \tag{1.6}$$

and

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - 1\hat{\beta})^T \mathbf{R}_D^{-1}(\mathbf{y} - 1\hat{\beta}). \tag{1.7}$$

The quantity to be optimised is a function of only the correlation parameters and the data.

Given the correlation parameters, [In practice, these parameters are not given but are estimated. For an update see Kleijnen and Van Beers (2003)] the next step is to build the best linear unbiased predictor (BLUP), $\hat{y}(\mathbf{x})$, of $y(\mathbf{x})$. The BLUP for an untried $\mathbf{x}$ is

$$\hat{y}(\mathbf{x}) = \hat{\beta} + \mathbf{r}^T(\mathbf{x})\mathbf{R}_D^{-1}(\mathbf{y} - 1\hat{\beta}) \tag{1.8}$$

where

$$\mathbf{r}(\mathbf{x}) = [R(\mathbf{x}_1, \mathbf{x}), \ldots, R(\mathbf{x}_n, \mathbf{x})]^T \tag{1.9}$$

is the vector of correlations between the zs at the design points $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ and at an untried input $\mathbf{x}$. Since $R(\mathbf{x}, \mathbf{x}) = 1$, the predictor will interpolate the data points, as it should if the data are without random error (Welch, *et al.*, 1992).

The prediction error for $\hat{y}(\mathbf{x})$ can be presented on the basis of the model considered in this thesis as

$$\mathrm{MSE}[\hat{y}(\mathbf{x})] = \sigma^2 \left[1 - \begin{pmatrix} 1 & \mathbf{r}^T(\mathbf{x}) \end{pmatrix} \begin{pmatrix} 0 & 1^T \\ 1 & \mathbf{R}_D \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mathbf{r}(\mathbf{x}) \end{pmatrix} \right]. \tag{1.10}$$

A number of design criteria were considered with the objective of choosing a design that predicts the response well at untried inputs in the experimental region. In the Integrated Mean Square criterion the objective is to minimize

$$\int_{\mathbf{x} \in R} \mathrm{MSE}[\hat{y}(\mathbf{x})] d\mathbf{x}$$

where R is the correlation matrix, $\hat{y}(\mathbf{x})$ is the predictor and MSE stands for mean square error. In the maximum mean squared error criterion a design is chosen

to minimise the maximum MSE although this is much more computationally demanding. A third criterion is the maximization of expected posterior entropy (amount of information available to the given experimental region). The asymptotic connection between the maximization of expected posterior entropy criteria (as established by Johnson *et al.*, 1990) is the need to insure that the design will be effective even if the response variable is sensitive to only a few design variables. The goal of this is to find designs which offer a compromise between the entropy maximin criterion, and projective properties in each dimension of the response variable (Morris and Mitchell, 1995). An approach by Ye *et al.*, (2000) on such "compromise between computing effort and design optimality" is a significant improvement on constructing designs for computer experiment. The authors claimed that the proposed class of designs [Symmetric Latin hypercube Design (**SLHDs**)] has some advantages over the regular LHDs with respect to criteria such as entropy and minimum intersite distance. The authors also claimed that SLHDs are a good subset of LHDs with respect to both entropy and maximin distance criteria (for a comprehensive discussion see Ye *et al.*, 2000).

A number of interesting points were made in the discussion of the paper by Sacks *et al.* (1989b). Some of the discussants raised the possibility of using factorial or fractional factorial design as cheap and simple designs that are nearly optimal in many cases. In their reply the authors made the claim that suitably scaled half fractions are apparently optimal or close to optimal for the IMSE criterion for the model with only a constant term for the regression and $p = 2$. A number of the discussants also examined the correlation function and stochastic process models and called for additional work in this area.

A Bayesian approach was adopted by Currin *et al.* (1991). They considered product of linear correlation functions

$$R(d) = 1 - \tfrac{1}{\theta}|d|, \quad \tfrac{1}{2} < \theta < \infty$$

and

$$\begin{aligned} R(d) &= 1 - \tfrac{1}{\theta}|d|, \quad |d| < \theta \\ &= 0, \qquad\quad |d| \geq \theta. \end{aligned}$$

and also product cubic correlation functions

$$\begin{aligned} R(d) &= 1 - 6(\tfrac{d}{\theta})^2 + 6(\tfrac{|d|}{\theta})^3, & |d| < \tfrac{\theta}{2}, \\ &= 2(1 - \tfrac{|d|}{\theta})^3, & \tfrac{\theta}{2} \leq |d| < \theta, \\ &= 0, & |d| \geq \theta \end{aligned}$$

where $\theta \geq 0$ and

$$R(\mathbf{d}) = \prod_{j=1}^{k} R(d_j).$$

The prediction functions are respectively linear or cubic splines in every dimension (Venables and Ripley, 1995, page 250). The design used was that which minimises the posterior entropy (Shewry and Wynn, 1987).

The authors gave a computational algorithm for finding entropy-optimal design on multidimensional grids. The authors found, however, that for some of the examples considered and for some correlation parameters the 95 percent posterior probability intervals do not give adequate convergence of the true values at selected test sites.

Welch *et al.* (1992) extended previous results to consider larger numbers of predictors in the case where there are only a few active factors. The correlation function considered was

$$R(\mathbf{w}, \mathbf{x}) = \prod_{j=1}^{d} \exp(-\theta_j |w_j - x_j|^{p_j}).$$

Although full maximum likelihood could be used it would be numerically costly. Instead each of the θ_j is set equal to each other, similarly for the p_j. The first stage is to maximize the likelihood based on the common values of θ and p. Then a stepwise procedure is used so that a θ_j and p_j is introduced for that variable for which such a step most increases the likelihood. The procedure continues until giving any of the remaining factors their own values of θ_j and p_j does not make a large difference relative to the previous stage.

Welch *et al.* (1992) note that there are numerical difficulties in maximising the likelihood and that it is difficult to obtain a global maximum. Different optima are sometimes obtained using different implementations of the same algorithm.

Good starting values are also obviously an important part of the MLE calculation. There is very little literature on how to get good starting values. One exception is Owen (1994a) who gave a method for estimating $\theta_1, \ldots, \theta_d$ when $p = 1$ which could be used as starting values in the maximum likelihood estimation.

A number of examples were presented by Welch *et al.* (1992). For one example involving only two variables a 50 run Latin Hypercube was shown to give good results. However, a 30 run and 40 run design does not identify a reasonable model. No guidelines for the appropriate sample size are given. In the second example a 50 run Latin Hypercube was successful in identifying five significant factors. The number of runs required for a computer experiment remains a key open question and is one that this research aims to tackle.

1.6 Thesis Outline

The thesis consists of seven chapters. The first chapter is a general introduction and Literature Review of the area of computer experiments.

In the second chapter a simple computer model will be introduced for the purpose of describing exactly what a computer model is, how it can be used and what problems will be addressed in the thesis. The approach of modeling the deterministic model as a stochastic process will be described. Focus will be on important practical issues that have not received much attention in the literature such as starting values, parameterization and generating plots for interpretation.

The third chapter will focus on the effect of sample size (the number of computer experiments) on the precision of the predictions based on the fitted stochastic model, particularly in the case when a Latin Hypercube design is used.

The fourth chapter will examine how to add extra runs to a computer experiment in order to improve the precision of the predictions. Two approaches will be examined, one adding one point at a time while the other adds groups of points at a time.

In the fifth chapter a new approach to addressing some of the issues of computer experiments will be developed. In this approach the structure of computer

experiments is used to study computer experiments themselves. This is likely to give results that are more general than the specific results developed in Chapters 3 and 4.

Many computer models give the response as a function of time. However most of the literature on computer experiments focused on each run of the code giving one response value. In Chapter 6 methods will be developed to efficiently analyse the multivariate data generated by a computer code.

Finally Chapter 7 will give a discussion of the results and point out areas for more research.

Chapter 2

Analysis of a Simple Computer Model

2.1 Introduction

Since this thesis focuses on a number of practical issues relating to the design and analysis of computer experiments, it is of value here to describe one model in detail. This material will be used in later chapters. The particular example selected is Available Safe Egress Time (ASET-B), a computer model predicting the nature of a fire in a single room, presented by Walton (1985).

A Latin Hypercube Design (LHD) is used to choose input factors for the ASET-B program. Based on this LHD, responses ($\mathbf{y}$) from the model are generated to form a computer experiment. The responses are modeled as the realisation of a stochastic process, following the work of Sacks *et al.* (1989a). Maximum Likelihood Estimates (MLE) of the parameters are generated and these estimates used to make predictions at untried inputs. The prediction can be made using the Best Linear Unbiased Predictor (BLUP), a methodology introduced by Henderson (1975b) and Goldberger (1962). A graphical interpretation of the results is presented.

2.2 Deterministic Fire Models

A stochastic process involves chance or uncertainty. In a deterministic world everything is assumed certain. Deterministic fire models attempt to represent mathematically the processes occurring in a compartment fire based on the laws of physics and chemistry. These models are also referred to as room fire models, computer fire models, or mathematical fire models. Ideally, they are such that discrete changes in any physical parameter can be evaluated in terms of the effect on fire hazard. While no such ideal exists in practice, a number of computer models are available that provide a reasonable amount of selected fire effects (Cooper and Forney, 1990).

Computer models have been used for some time in the design and analysis of fire protection hardware. The use of computer models, commonly known as design programs, has become the industry's standard method for designing water supply and automated sprinkler systems. These programs perform a large number of tedious and lengthy calculations and provide the user with accurate, cost-optimised designs in a fraction of the time that would be required for manual procedures.

In addition to the design of fire protection hardware, computer models may also be used to help evaluate the effects of fire on both people and property. The models can provide a fast and more accurate estimate of the impact of a fire and help establish the measures needed to prevent or control it. While manual calculation methods provide good estimates of specific fire effects (eg., prediction of time to flash over), they are not well suited for comprehensive analysis involving the time-dependent interactions of multiple physical and chemical processes present in developing fires.

2.2.1 The ASET-B Fire Model

Fire is a complex phenomenon and a number of computer models have been developed that reflect this complexity, for use by scientists and engineers. One of the earliest models was the ASET (Available Safe Egress Time) mathematical model

written in FORTRAN by Cooper (1980). Later, Walton (1985) implemented the model in Basic as ASET-B incorporating simpler numerical techniques to solve the differential equations involved.

ASET-B is a personal computer program for predicting the fire environment in a single room enclosure with all doors, windows and vents closed except for a small leak at floor level. This leak prevents the pressure from increasing in the room. A fire starts at some point below the ceiling and releases energy and the products of combustion. The rate at which these are released is likely to change with time. The hot products of combustion form a plume which, due to buoyancy, rises. As it does so, it draws in to the room cool air which decreases the plume's temperature and increases its volume flow rate. When the plume reaches the ceiling it spreads out and forms a hot gas layer which descends with time as the plume's gases continue to flow into it. There is a relatively sharp interface between the hot upper layer and the air in the lower part of the room which, in the ASET-B model, is considered to be at ambient temperature. The only interchange between the air in the lower part of the room and the hot upper layer is through the plume.

ASET-B solves several differential equations using a simpler numerical technique than in the original ASET program. ASET-B requires as inputs the height and area of the room, the elevation of the fire above the floor, a heat loss factor (the fraction of the heat released by the fire that is lost to the bounding surfaces of the enclosure) and a fire specified in terms of heat release rate which depends on the nature of the combustion material. For this study I have used the rate of release for a 'semi-universal fire', corresponding to a "fuel package consisting of a polyurethane mattress with sheets, fuels similar to wood cribs and polyurethane on pallets, and commodities in paper cartons stacked on pallets" (Birk 1991, page 86). The program predicts the thickness and the temperature of the hot smoke layer as a function of time. A simple illustration of fire-in-enclosure flow dynamics for an "unvented" enclosure and the basic fire phenomena are presented in Figure 2.1.

The response (**y**) was taken as the time it takes for the height of the smoke

layer to be at 5 ft (head height). This manipulation was carried out in order to make the output univariate. The analysis of profiles over time is presented in chapter 6.

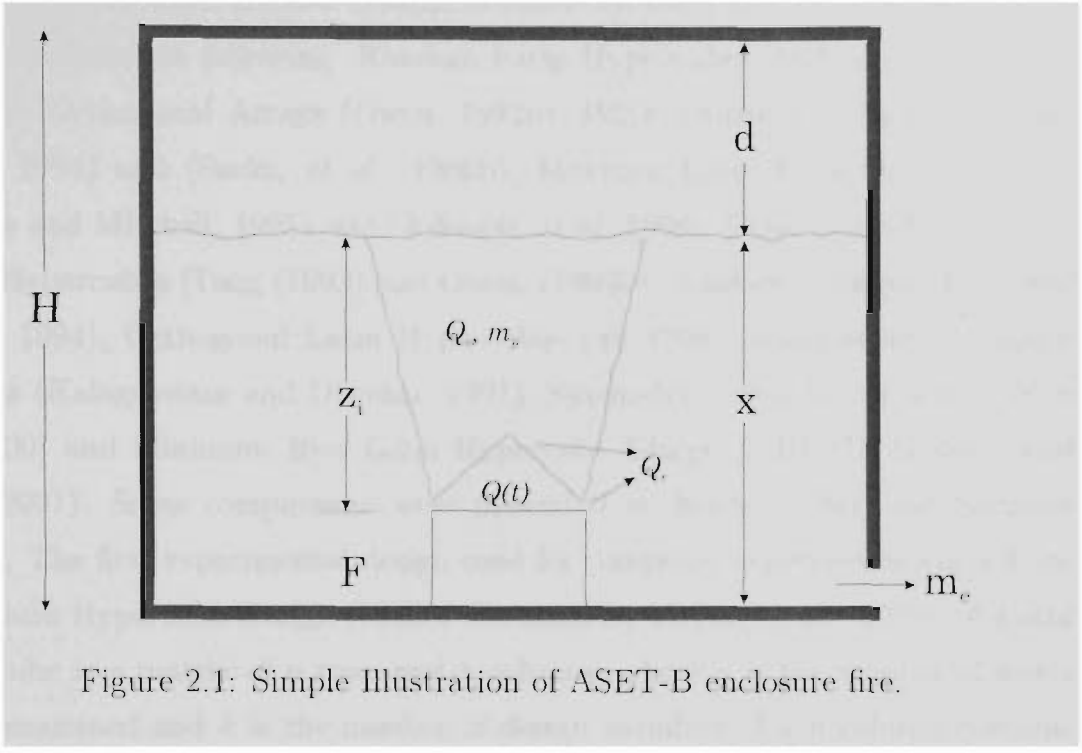

Figure 2.1: Simple Illustration of ASET-B enclosure fire.

F	–	Height of base of fire
H	–	Height of room
$\mathbf{m_e}$	–	Mass flow rate leaving crack like vent
$\mathbf{m_p}$	–	Plume mass flow rate
$\mathbf{Q_c}$	–	Convective energy release rate
$\mathbf{Q_r}$	–	Radiative energy release rate
$\mathbf{Q_{(t)}}$	–	Heat release rate at time **(t)**
X	–	Height of interface above floor
$\mathbf{Z_i}$	–	Interface height above fuel surface.

Reprinted from [Birk (1991, p. 35)], with permission from Technomic Publishing Co., Inc., (1991)©.

2.3 Experimental Design

2.3.1 Latin Hypercube Designs

There are many experimental designs available for computer experiments. These designs include the following: Random Latin Hypercubes [McKay, *et al.* 1979], Random Orthogonal Arrays (Owen, 1992b), IMSE Optimal Latin Hypercubes (Park, 1994) and (Sacks, *et al.* 1989b), Maximin Latin Hypercubes [MmLh] (Morris and Mitchell, 1995) and (Johnson, *et al.* 1990), Orthogonal-Array based Latin Hypercubes [Tang (1993) and Owen, (1992b)], Uniform Designs (Fang and Wang, 1994), Orthogonal Latin Hypercubes (Ye 1998), Hammersley Sequence Designs (Kalagnanam and Diwekar, 1997), Symmetric Latin Hypercubes (Ye *et al.* 2000) and Minimum Bias Latin Hypercube Design [MBLHD] (Palmer and Tsui, 2001). Some comparisons were presented by Koeler (1990) and Simpson (1998). The first experimental design used for computer experiments was a Random Latin Hypercube Design (LHD), discussed by McKay *et al.* (1979). A Latin hypercube is a matrix of n rows and k columns where n is the number of levels being examined and k is the number of design variables. Each column contains the levels 1, 2,..., n, randomly permuted, and the k columns are matched at random to form the Latin hypercube. By their nature, Latin hypercubes are quite easy to generate because they require only a random permutation of n levels in each column of the design matrix. The big advantage of Latin hypercube designs is that they ensure stratified sampling, i.e., each of the input variables is sampled at n levels. Thus, when a Latin hypercube is projected or collapsed into a single dimension, n distinct levels are obtained. This is extremely beneficial for deterministic computer experiments since the Latin hypercube points do not overlap, minimizing any information loss. The main attraction of these designs is that they have good one-dimensional projective properties which ensure that there is little redundancy of design points when some of the factors have a relatively negligible effect, called effect sparsity (Butler, 2001). "The use of Latin hypercubes does not put any serious restrictions on the applicability of the importance sampling methods and the benefits of Latin hypercubes can be added to

the benefits of elaborate and efficient importance sampling strategies. Therefore, Latin hypercube sampling has the qualifications to become a widely employed tool in reliability analysis." (Olsson *et al.*, 2003).

A LHD gives an evenly distributed projection into each of the input factors. With this choice in a sample of size N the j^{th} observation for the i^{th} variable is given as

$$X_i^j = \left(\frac{\text{Max}_i - \text{Min}_i}{2}\right) x_i^j + \left(\frac{\text{Min}_i + \text{Max}_i}{2}\right) \quad i = 1, \ldots, d; j = 1, \ldots, N$$

where Max_i is the Maximum of X_i, Min_i is the Minimum of X_i,

$$x_i^j = \frac{2\pi_j(i) - N - 1}{(N-1)}$$

and $\pi_j(i)$ is the j^{th} observation of a random permutation of the integers $\mathbf{1}, \ldots, N$, $\pi(j) = (\pi_j(\mathbf{1}), \ldots, \pi_j(N))$ and the d random permutations $\pi(\mathbf{1}), \ldots, \pi(d)$ corresponding to the d input factors are mutually independent. Note that -1 to 1 has been chosen as the range for the coded input factors, x_i, although some authors (Sacks *et al.* 1989b) prefer a range of 0 to 1.

A simple LHD ($N = 11$), for a two dimensional (x_1 and x_2) case is illustrated in Figure 2.2 and shows that each of those components is represented in a fully stratified manner. Note that each component is sampled uniformly on the interval $[-1, 1]$.

2.3.2 Application to ASET-B Computer Experiment

The first stages of a computer experiment involve selecting the input variables and the ranges over which they will be explored. For the ASET-B model the inputs were taken to be the Heat Loss Fraction, the Fire Height, the Room Ceiling Height and the Room Floor Area giving a four dimensional configuration. The ranges of the variables are given in Table 2.1.

As indicated previously the input variables $X_1, \ldots, X_4$, were coded as $x_1, \ldots, x_4$; where the x_i have a range of -1 to 1.

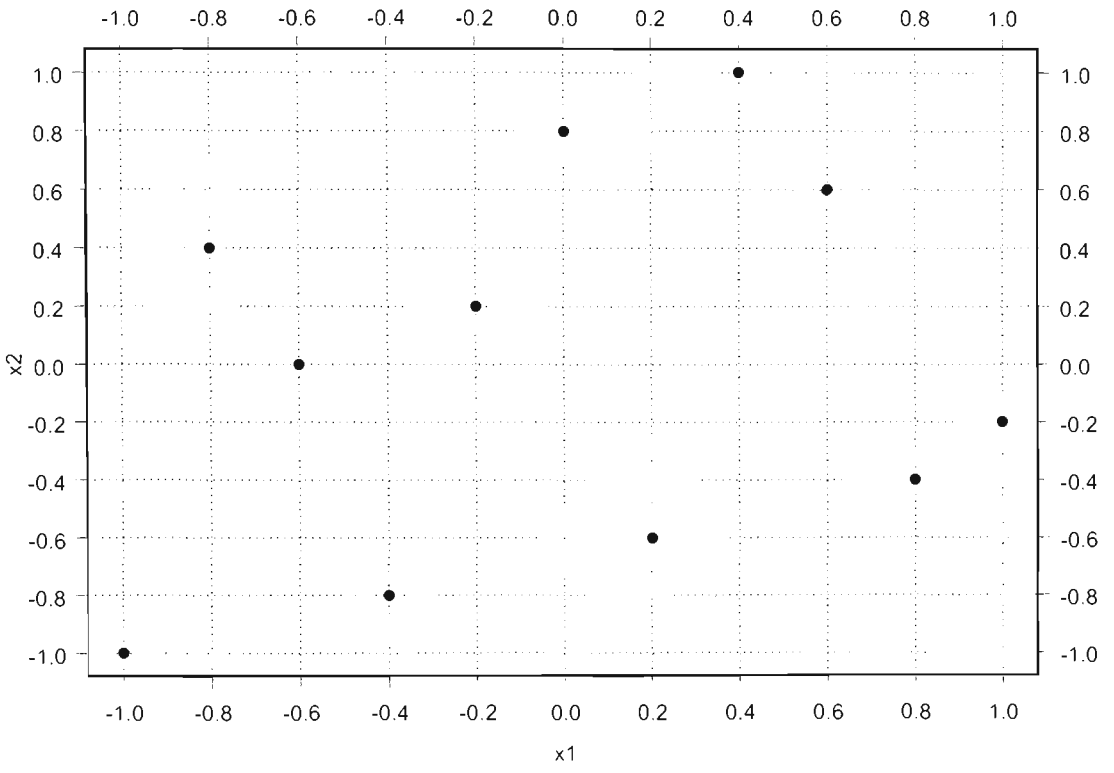

Figure 2.2: Projection Properties of a LHD with 11 runs.

The number of runs required remains an open question for computer experiments. Welch *et al.* (1996) suggest, as a guideline, that the number of runs in a computer experiment should be chosen to be 10 times the number of active inputs, which would lead to $N = 40$ runs for this example if all four factors turn out to be active. To be conservative $N = 50$ runs was used. More on sample size considerations will be discussed in Chapter 3.

The actual input variables and response (Egress Time), generated from the ASET-B program, are given in Table 2.2. Table 2.3 and a pictorial representation of the design is given in Figure 2.3. The egress time for each run of the LHD was calculated using linear interpolation by assuming that the height of the smoke layer is 5ft (head height).

Variable	Variable name	Minimum	Maximum
X_1	Heat Loss Fraction	0.6	0.9
X_2	Fire Height (ft)	1.0	3.0
X_3	Room Ceiling Height (ft)	8.0	12.0
X_4	Room Floor Area (sq. ft)	81.0	256.0

Table 2.1: Input Variables for ASET-B Fire Model

2.4 Modelling

2.4.1 Summary of the Approach used by Sacks *et al.*

Modelling for the ASET-B responses were carried out as a realization of a stochastic process, following the methodology developed by Sacks *et al.* (1989b).

The response $y(\mathbf{x})$ is assumed to follow Equation 1.4 where the systematic component $z(\mathbf{x})$ is modelled as a realisation of a Gaussian stochastic process in which the covariance function of $z(\mathbf{x})$ relates to the smoothness of the response. The covariance of the responses to two d-dimensional inputs $\mathbf{t} = (t_1, \ldots, t_d)$ and $\mathbf{u} = (u_1, \ldots, u_d)$ is given by Equation 1.1 and Equation 1.2

2.4.2 Maximum Likelihood Estimation

Sacks *et al.* (1989b) show that obtaining the maximum likelihood estimators of $\theta_1, \ldots, \theta_4, p_1, \ldots, p_4$, β and σ^2 reduces to numerically optimising Equation 1.5, with $\hat{\beta}$ given by Equation 1.6 and $\hat{\sigma}^2$ given by Equation 1.7. Note that the quantity to be optimised is a function of only the correlation parameters, (θ_j and p_j) and the data.

Starting Values

After assigning the inputs to LHD for different sizes of N, the next step is to choose starting values for the Maximum Likelihood Estimation. Calculation of the maximum likelihood estimates depends on good starting values. To study a suitable distribution for starting values, some values listed in the literature for θ and p were reviewed. Table 2.4 summarise the values for $\hat{\theta}$ and $\hat{p}$ as listed.

Run No.	$49x_1$	$49x_2$	$49x_3$	$49x_4$	**(Egress Time)**
1	−37	−11	−31	−27	26.00
2	23	−31	−19	−19	58.33
3	37	5	−41	−49	36.25
4	−17	11	25	25	48.75
5	−47	−25	−3	35	40.00
6	49	39	27	31	25.83
7	11	−9	43	45	55.00
8	17	−1	29	5	61.67
9	−19	31	37	−41	55.00
10	−25	43	−45	47	34.00
11	−13	−21	−35	−25	43.00
12	47	21	21	33	63.33
13	−45	−47	−15	41	73.33
14	−11	15	−33	−47	43.75
15	−49	29	31	−9	33.00
16	−21	9	13	13	61.67
17	13	47	−21	−39	35.00
18	−29	−37	−23	−29	48.75
19	19	−45	−25	−43	61.67
20	−9	−27	35	1	56.67
21	43	13	23	3	48.33
22	−1	−15	−47	29	33.00
23	−3	−41	3	−15	58.75
24	−7	35	9	−31	47.50
25	−39	−43	5	37	56.25

Table 2.2: Scaled LHD points for ASET-B Input Variables. [Runs 1 to 25 of $N = 50$] and corresponding Egress Times. **Coding scheme:** scaled $[-1, +1]$ input variables have been multiplied by 49.

Run No.	$49x_1$	$49x_2$	$49x_3$	$49x_4$	**(Egress Time)**
26	5	23	47	49	27.00
27	39	33	−49	19	46.67
28	31	−7	−29	−3	57.50
29	−41	−19	19	−5	51.25
30	−23	−33	−7	−7	30.00
31	15	1	−17	15	77.50
32	−31	−35	33	43	48.33
33	41	41	−27	−45	25.83
34	29	−5	7	−33	61.67
35	9	27	−37	−1	46.67
36	−33	49	17	−23	37.50
37	−43	−17	39	−37	34.00
38	−27	7	−9	11	35.00
39	−35	−13	−1	7	37.00
40	7	37	45	17	60.00
41	35	17	−13	−21	31.25
42	25	45	49	−13	42.50
43	45	−23	−5	−17	57.50
44	3	25	1	−11	37.00
45	−5	3	41	27	35.00
46	27	−3	−43	39	45.00
47	21	−49	11	21	37.50
48	−15	−29	−39	23	38.75
49	1	−39	−11	9	52.50
50	33	19	15	−35	63.33

Table 2.3: Scaled LHD points for ASET-B Input Variables. [Runs 26 to 50 of $N = 50$] and corresponding Egress Times. **Coding scheme:** scaled $[-1, +1]$ input variables have been multiplied by 49.

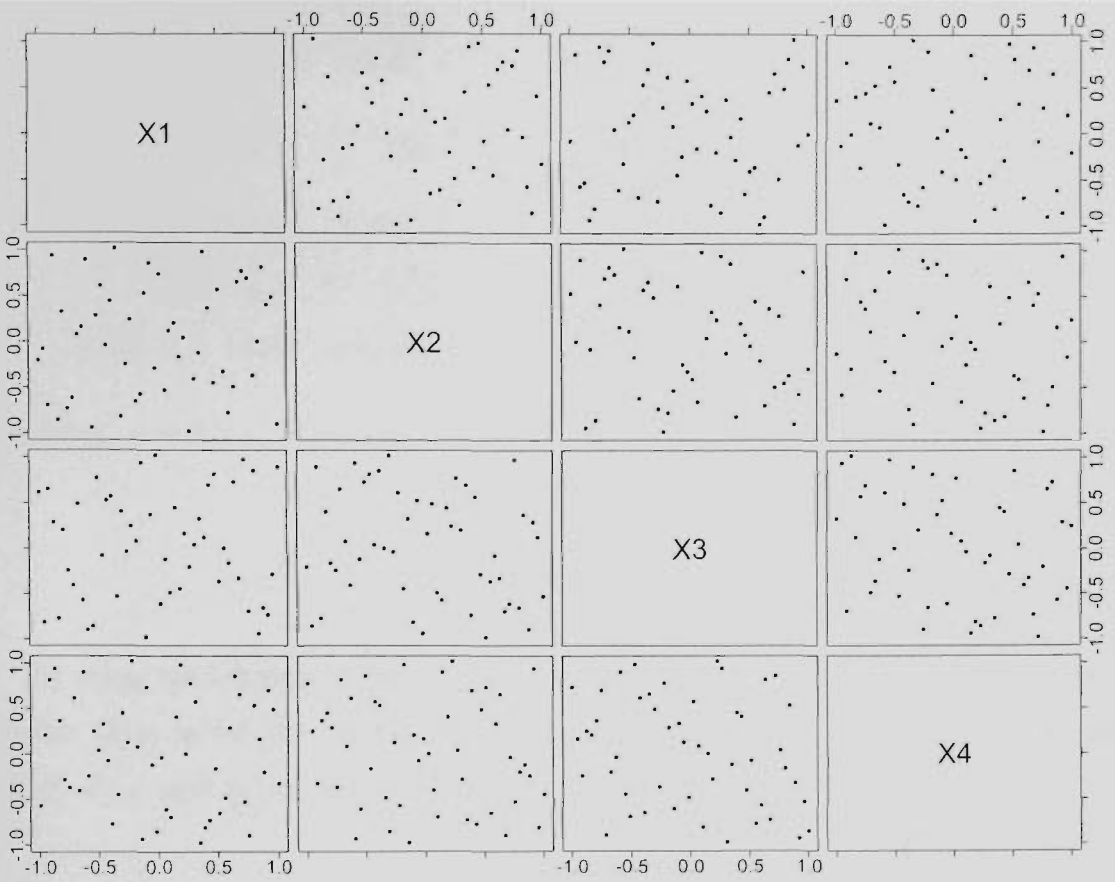

Figure 2.3: Projection Properties of a LHD with 50 runs.

According to the indicated magnitudes of θ's and p's: the θs lie between 0 and 2 and the ps will also be between 0 and 2.

Based on this evidence, the method adopted was to generate 10 random sets of starting values and for each set to calculate the maximum likelihood estimates. The random starts for the parameters p_j, $j = 1, \ldots, d$, were generated from a uniform distribution on $[0, 2]$, while the random starts for the θ_j, $j = 1, \ldots, d$, were generated from an exponential distribution with mean 1. Of course other distributions could be used.

Calculating the maximum likelihood estimates is a constrained optimisation problem since $0 \leq p_j \leq 2$, $j = 1 \ldots d$ and $0 < \theta_j$ $j = 1 \ldots d$.

More success resulted by turning the constrained optimisation problem involving $\theta_1, \ldots, \theta_d, p_1, \ldots, p_d$ into an unconstrained optimisation involving parameters

Citation	Case	θ_{min}	θ_{max}	p_{min}	p_{max}
Sacks *et al.* (1989b)	1	0.00	1.970	1.61	2.00
Koehler (1990)	2	0.00	1.800	1.00	2.00
	3	0.00	1.850	2.00	2.00
Welch *et al.* (1992)	4	0.00	0.036	1.70	2.00
	5	0.00	0.970	1.61	2.00
Booker (1998)	6	−0.4	0.4	0.00	0.25
Chang *et al.* (1999)	7	0.00	2.00	0.00	2.00

Table 2.4: Some suggested values for θs and ps from selected cases.

ϕ_j and q_j, $j = 1, \ldots, d$, where

$$\begin{aligned} \phi_j &= \pm\sqrt{\theta_j} \\ q_j &= \ln\left(\sqrt{\frac{2}{p_j}}\right). \end{aligned}$$

By using the likelihood based on the ϕ_j and q_j, convergence was achieved much quicker than using the likelihood based on the θ_j and p_j. Once the converged values of ϕ_j and q_j are obtained the estimated values of θ_j and p_j are given by

$$\begin{aligned} \hat{\theta}_j &= \hat{\phi}_j^2 \\ \hat{p}_j &= \frac{2}{(\exp \hat{q}_j)^2}. \end{aligned}$$

As a preliminary experiment two sets of ten random starting values for the ASET-B model were generated. The results for the first set of starting values are given in Table 2.5 and Table 2.6. All ten starting values give the same maximum likelihood estimates (−loglikelihood = 12.12207). The results for the second set of ten starting values are given in Table 2.7 and Table 2.8. In contrast to the first set, the second set gave four different likelihood modes. The mode that occurs most often appears to be the maximum likelihood and gives the same loglikelihood as in the first case.

2.4.3 Application to ASET-B Computer Experiment

The estimates of the parameters for the data in Table 2.2 and Table 2.3 obtained using maximum likelihood are given in Table 2.9. Ten random starting values

Random Start	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
1	0.0000000	0.1512159	0.0008050	0.2441341
2	0.0000000	0.1511380	0.0008048	0.2444890
3	0.0000000	0.1512156	0.0008011	0.2445132
4	0.0000000	0.1510373	0.0008084	0.2484577
5	0.0000000	0.1510374	0.0008084	0.2484576
6	0.0000000	0.1512159	0.0007999	0.2451331
7	0.0000000	0.1510375	0.0008084	0.2484577
8	0.0000000	0.1510374	0.0008084	0.2484576
9	0.0000000	0.1510374	0.0008084	0.2484576
10	0.0000000	0.1512157	0.0007998	0.2451331

Table 2.5: Estimates of $\theta_1, \ldots, \theta_4$ from 10 Random Starts with one Maximum (12.12207).

Random Start	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$
1	1.499745	2.000000	0.367233	2.000000
2	1.512299	2.000000	0.369722	2.000000
3	1.520238	1.999118	0.369337	2.000000
4	1.574930	2.000000	0.379776	2.000000
5	1.352069	2.000000	0.379776	2.000000
6	1.369863	1.994118	0.368021	2.000000
7	1.415365	2.000000	0.379778	2.000000
8	1.974627	2.000000	0.379775	2.000000
9	1.618871	2.000000	0.379775	2.000000
10	1.938354	1.994118	0.367884	2.000000

Table 2.6: Estimates of $p_1, \ldots, p_4$ from 10 Random Starts with one Maximum (12.12207).

were used and each set converged to the same mode.

2.5 Prediction

2.5.1 Prediction for Untried Inputs

Given the estimated parameters, prediction at untried inputs can be made using BLUP (see, for example, Robinson, 1991). The prediction at $\mathbf{x}$ is given by Equation 1.8 and Equation 1.9.

Random Start	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	MLE
1**	0.0000000	0.1362153	0.0000000	0.2051321	12.33658
2	0.0000000	0.1510374	0.0008084	0.2484577	12.12207
3	0.0000000	0.1510375	0.0008084	0.2484578	12.12207
4	0.0000000	0.1510374	0.0008084	0.2484577	12.12207
5	0.0000000	0.1510374	0.0008084	0.2484576	12.12207
6*	0.0000000	0.1488380	0.0000000	0.2494890	12.17769
7	0.0000000	0.1510374	0.0008084	0.2484577	12.12207
8**	0.0000000	0.1362151	0.0000000	0.2051327	12.33658
9	0.0000000	0.1510371	0.0008084	0.2484573	12.12207
10***	0.0000000	0.1642061	0.0000000	0.2690531	12.55905

Table 2.7: Estimates of $\theta_1, \ldots, \theta_4$ from 10 Random Starts with Different Maximum (MLE).

Random Start	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	MLE
1**	1.940002	1.949119	1.994774	2.000000	12.33658
2	1.640228	2.000000	0.379775	2.000000	12.12207
3	1.883667	2.000000	0.379776	2.000000	12.12207
4	1.599643	2.000000	0.379779	2.000000	12.12207
5	1.406527	2.000000	0.379775	2.000000	12.12207
6*	0.000000	2.000000	1.896527	2.000000	12.17769
7	1.902768	2.000000	0.379775	2.000000	12.12207
8**	0.216133	1.949118	1.949424	2.000000	12.33658
9	1.101159	2.000000	0.379773	2.000000	12.12207
10***	1.451325	2.000000	1.889690	1.957572	12.55905

Table 2.8: Estimates of $p_1, \ldots, p_4$ from 10 Random Starts with Different Maximum (MLE).

j	1	2	3	4
$\hat{\theta}_j$	0.00934	0.02041	0.01953	0.02849
$\hat{p}_j$	1.98051	1.89641	1.99708	2.00000
$\hat{\beta}$	48.3526			
$\hat{\sigma}^2$	123.556			

Table 2.9: Estimates of Parameters for the ASET-B Computer Model.

2.5.2 Prediction Error

For a prediction to be useful it should be supplemented by a measure of its precision. A number of different measures have been introduced for computer experiments. Their utility has been reviewed by Sacks *et al.* (1989a). The most important measures are:

Empirical Mean Square Error

$$\text{EMSE} = \left\{ \frac{1}{N} \sum_{N} [\hat{y}(\mathbf{x}) - y(\mathbf{x})]^2 \right\} \tag{2.1}$$

where $\mathbf{x}$ is a set of N randomly selected points over the experimental region $\mathcal{R}$.

A related measure is

Empirical Root Mean Square Error

$$\text{ERMSE} = \left\{ \frac{1}{N} \sum_{N} [\hat{y}(\mathbf{x}) - y(\mathbf{x})]^2 \right\}^{\frac{1}{2}} \tag{2.2}$$

Mean Square Error at a point x

$$\text{MSE}[\hat{y}(\mathbf{x})] = \sigma^2 \left[1 - \begin{pmatrix} 1 & \mathbf{r}^T(\mathbf{x}) \end{pmatrix} \begin{pmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{R}_D \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mathbf{r}(\mathbf{x}) \end{pmatrix} \right] \tag{2.3}$$

Maximum Mean Square Error

$$\text{Max}_{\mathbf{x} \in \mathcal{R}} \text{MSE}(\hat{y}(\mathbf{x})) \tag{2.4}$$

Integrated Mean Square Error

$$\int_{\mathbf{x} \in \mathcal{R}} \text{MSE}(\hat{y}(\mathbf{x})). \tag{2.5}$$

2.5.3 Application to ASET-B Computer Experiment

To see the usefulness of the predictions the ASET-B model was run for 100 random points over the design range and predictions made based on the fitted computer model. Figure 2.4 shows that the predictions match the actual responses from ASET-B quite closely.

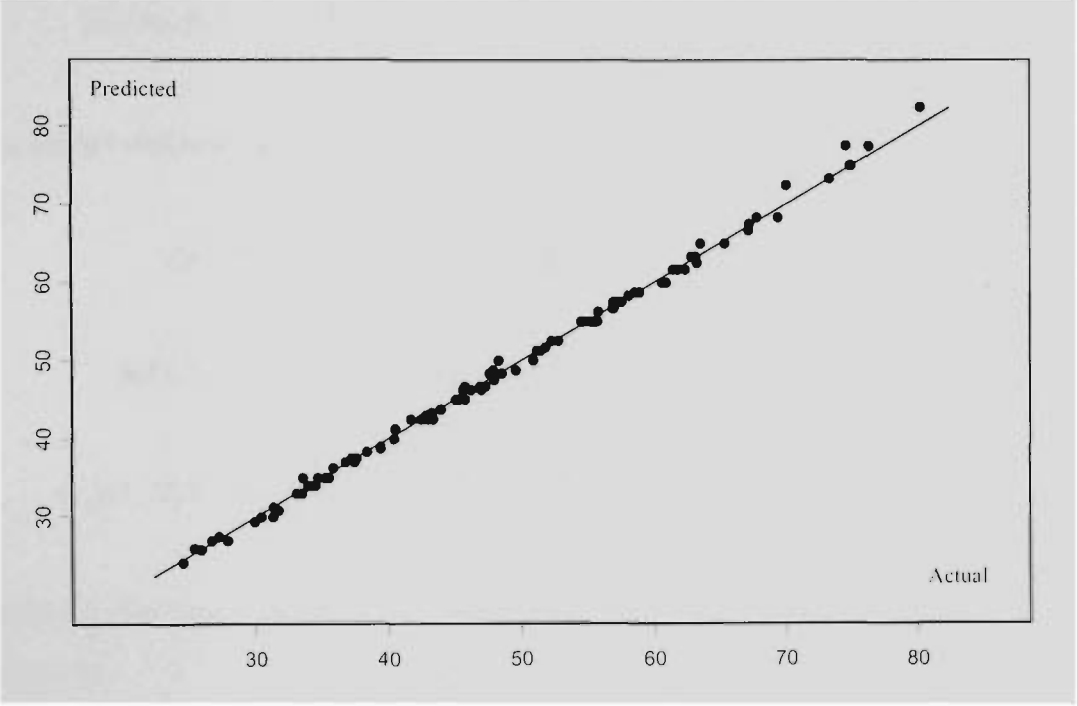

Figure 2.4: Accuracy of Prediction for Egress Time

2.6 Interpretation of Results

2.6.1 Analysis of Main Effects and Interactions

This section of the chapter demonstrates how the main effects and interaction plots which have been used by a number of authors can be efficiently calculated. An example of this is given below.

Sacks *et al.* (1992) defined, for a computer model with input range $\mathbf{u} \in [0,1]^d$, the mean, main effects and interaction effects as

$$\begin{aligned}
\mu_0 &= \int_{[0,1]^d} \cdots \int y(\mathbf{u}) \prod_{h=1}^{d} du_h \\
\mu_i(u_i) &= \int_{[0,1]^d} \cdots \int y(\mathbf{u}) \prod_{h \neq i} du_h - \mu_0 \\
\mu_{i,j}(u_i, u_j) &= \int_{[0,1]^d} \cdots \int y(\mathbf{u}) \prod_{h \neq i,j} du_h - \mu_i(u_i) - \mu_j(u_j) - \mu_0.
\end{aligned}$$

For a model defined on $\mathbf{z} \in [-1,1]^d$ the definitions become

$$\begin{aligned}
\mu_0 &= \frac{1}{2^d} \int_{[-1,1]^d} \cdots \int y(\mathbf{z}) \prod_{h=1}^{d} dz_h \\
\mu_i(z_i) &= \frac{1}{2^{d-1}} \int_{[-1,1]^d} \cdots \int y(\mathbf{z}) \prod_{h \neq i} dz_h - \mu_0 \\
\mu_{i,j}(z_i, z_j) &= \frac{1}{2^{d-2}} \int_{[-1,1]^d} \cdots \int y(\mathbf{z}) \prod_{h \neq i,j} dz_h - \mu_i(z_i) - \mu_j(z_j) - \mu_0.
\end{aligned}$$

In practice, the true effects are estimated by replacing $y(\mathbf{z})$ by $\hat{y}(\mathbf{z})$ in the above expressions.

Since the BLUP estimator of y is given by Equation 1.8 then the estimator $\hat{y}$ can be re-written as:

$$\hat{y}(\mathbf{z}) = \hat{\beta} + \mathbf{r}^T(\mathbf{z})\mathbf{w}$$

where

$$\mathbf{w} = \mathbf{R}_D^{-1}(\mathbf{y} - \mathbf{1}\hat{\beta})$$

and $\mathbf{r}(\mathbf{z})$ is given by Equation 1.9. Hence $\hat{y}(\mathbf{z})$ can be evaluated as

$$\hat{y}(\mathbf{z}) = \hat{\beta} + w_1 R(\mathbf{x}_1, \mathbf{z}) + w_2 R(\mathbf{x}_2, \mathbf{z}) + \ldots + w_n R(\mathbf{x}_n, \mathbf{z})$$

for an arbitrary input point $\mathbf{z}$.

For a 1 dimensional case, as here, consider

$$\begin{aligned}
\frac{1}{2^4}\underset{[-1,1]^4}{\iiiint}\hat{y}(\mathbf{z})\,dz_1\,dz_2\,dz_3\,dz_4 &= \frac{1}{2^4}\underset{[-1,1]^4}{\iiiint}\left[\hat{\beta}+\sum_{k=1}^{n}w_k R(\mathbf{x}_k,\mathbf{z})\right]dz_1\,dz_2\,dz_3\,dz_4 \\
&= \frac{1}{2^4}\left[\hat{\beta}\underset{[-1,1]^4}{\iiiint}dz_1\,dz_2\,dz_3\,dz_4\right. \\
&\quad \left.+\sum_{k=1}^{n}w_k\underset{[-1,1]^4}{\iiiint}R(\mathbf{x}_k,z)\,dz_1\,dz_2\,dz_3\,dz_4\right] \\
&= \hat{\beta}+\frac{1}{2^4}\sum_{k=1}^{n}w_k\underset{[-1,1]^4}{\iiiint}R(\mathbf{x}_k,z)\,dz_1\,dz_2\,dz_3\,dz_4 \\
&= \hat{\beta}+\frac{1}{2^4}\sum_{k=1}^{n}w_k\left[\prod_{h=1}^{4}\int_{-1}^{1}\exp\left(-\theta_h\left|x_{hk}-z_h\right|^{p_h}\right)dz_h\right].
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{1}{2^3}\underset{[-1,1]^3}{\iiint}\hat{y}(z)\,dz_2\,dz_3\,dz_4 &= \frac{1}{2^3}\underset{[-1,1]^3}{\iiint}\left[\hat{\beta}+\sum_{k=1}^{n}w_k R(\mathbf{x}_k,\mathbf{z})\right]dz_2\,dz_3\,dz_4 \\
&= \frac{1}{2^3}\left[\hat{\beta}\underset{[-1,1]^3}{\iiint}dz_2\,dz_3\,dz_4\right. \\
&\quad \left.+\sum_{k=1}^{n}w_k\underset{[-1,1]^3}{\iiint}R(\mathbf{x}_k,\mathbf{z})\,dz_2\,dz_3\,dz_4\right] \\
&= \hat{\beta}+\frac{1}{2^3}\left[\sum_{k=1}^{n}w_k\exp\left(-\theta_1\left|x_{k1}-z_1\right|^{p_1}\right)\right. \\
&\quad \left.\times\prod_{h=2}^{4}\int\exp\left(-\theta_h\left|x_{hk}-z_h\right|^{p_h}\right)dz_h\right].
\end{aligned}$$

Finally,

$$\begin{aligned}
\frac{1}{2^2}\iint_{[-1,1]^2} \hat{y}(z)\,dz_3\,dz_4 &= \frac{1}{2^2}\iint_{[-1,1]^2} \hat{\beta} + \sum_{k=1}^{n} w_k R(\mathbf{x}_k, \mathbf{z})\,dz_3\,dz_4 \\
&= \frac{1}{2^2}\left[\hat{\beta}\iint_{[-1,1]^2} dz_3\,dz_4 \right. \\
&\quad \left. + \sum_{k=1}^{n} w_k \iint_{[-1,1]^2} R(\mathbf{x}_k, \mathbf{z})\,dz_3\,dz_4\right] \\
&= \hat{\beta} + \frac{1}{2^2}\left[\sum_{k=1}^{n} w_k \exp\left(-\theta_1 |x_{k1} - z_1|^{p_1} - \theta_2 |x_{k2} - z_2|^{p_2}\right)\right. \\
&\quad \left. \times \prod_{h=3}^{4} \int_{-1}^{1} \exp\left(-\theta_h |x_{hk} - z_h|^{p_h}\right) dz_h\right].
\end{aligned}$$

Note that all these quantities only depend on one-dimensional integrals

$$\int_{-1}^{1} \exp\left(-\theta_h |x_{hk} - z_k|^{p_h}\right) dz_k$$

where x_{hk} is h^{th} component of the k^{th} point of the initial experiment and z_h is the h^{th} component of an arbitrary point. This is so since the integrand is separable in z_h.

In general

$$\mu_0 = \hat{\beta} + \frac{1}{2^d}\sum_{k=1}^{n} w_k \left[\prod_{j=1}^{d} \int_{-1}^{1} \exp\left(-\theta_j |x_{ij} - z_j|^{p_j}\right) dz_j\right] \tag{2.6}$$

$$\begin{aligned}
\mu_i(z_i) &= \hat{\beta} + \frac{1}{2^{d-1}}\left[\sum_{k=1}^{n} w_k \exp\left(-\theta_i |x_{kj} - z_j|^{p_j}\right)\right. \\
&\quad \left. \times \prod_{h\neq i} \int_{-1}^{1} \exp\left(-\theta_j |x_{ij} - z_j|^{p_h}\right) dz_j\right] - \mu_0
\end{aligned} \tag{2.7}$$

$$\begin{aligned}
\mu_{i,j}(z_i, z_j) &= \hat{\beta} + \frac{1}{2^{d-2}}\left[\sum_{k=1}^{n} w_k \exp\left(-\sum_{h=i,j} \theta_h |x_{kh} - z_h|^{p_h}\right)\right. \\
&\quad \left. \prod_{h\neq i,j} \int_{-1}^{1} \exp\left(-\theta_h |x_{kh} - z_h|^{p_h}\right) dz_h\right] - \mu_i(z_i) - \mu_j(z_j) - \mu_0.
\end{aligned} \tag{2.8}$$

2.6.2 Application to ASET-B Computer Experiment

Using the results of the previous section, estimates of the average over the experimental region, given by Equation 2.6, and the main effect of input factor x_i (averaged over the other factors), given by Equation 2.7, were calculated. Estimates were obtained by replacing $y(\mathbf{z})$ by $\hat{y}(\mathbf{z})$. The estimated value of μ_0 is 48.75. The main effect function is given in Figure 2.5, showing that the Egress Time increases as each of the input factors increases, with the most important factors over the ranges studied being Room Floor Area (D) and Fire Height (B). Heat Loss Fraction (A) and Room Ceiling Height (C) are less important in this model.

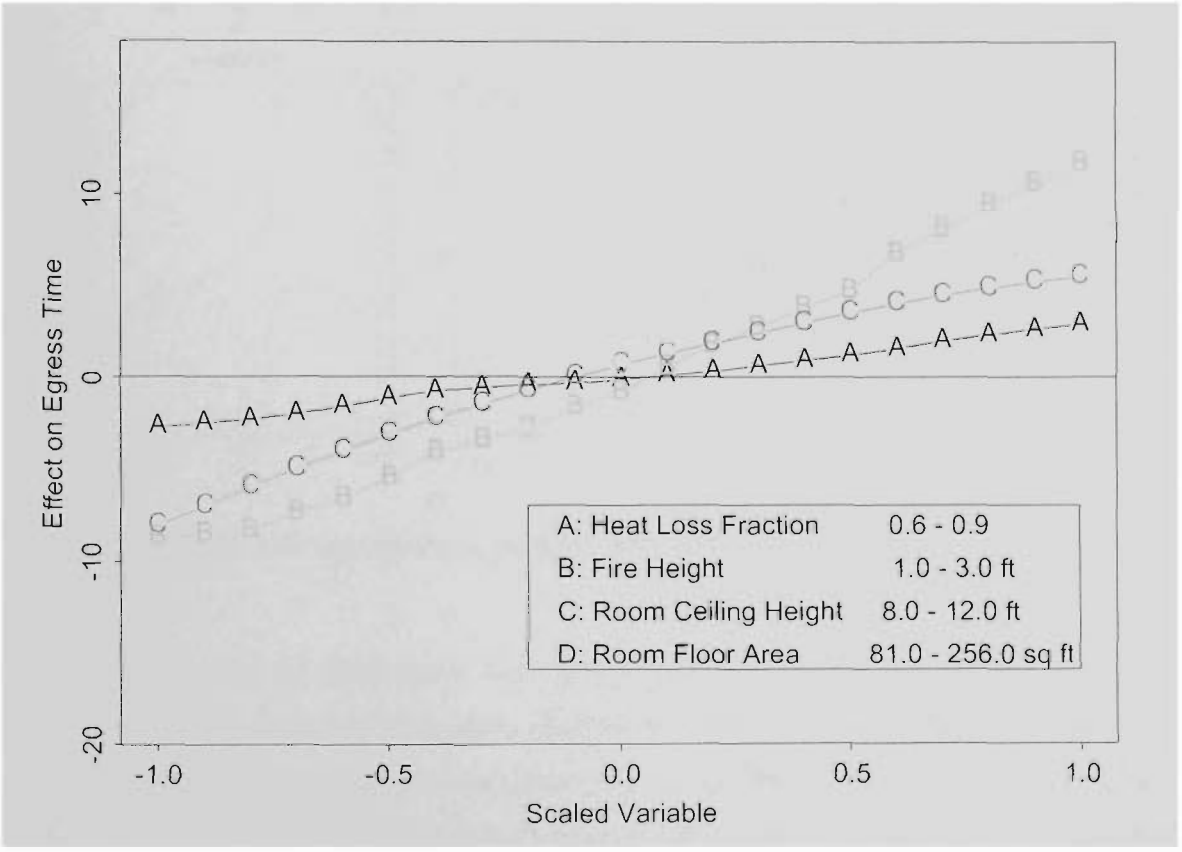

Figure 2.5: Main Effects plot for the ASET-B Computer Model.

Similarly the estimated interaction effects of z_i and z_j, given by Equation 2.8

were calculated, with estimates obtained by replacing $y(\mathbf{z})$ by $\hat{y}(\mathbf{z})$. Contours of the interaction functions are given in Figure 2.6. The deviations from additivity are quite small (since the interactions are small relative to main effects). The joint effects of each pair of factors are given in Figure 2.7.

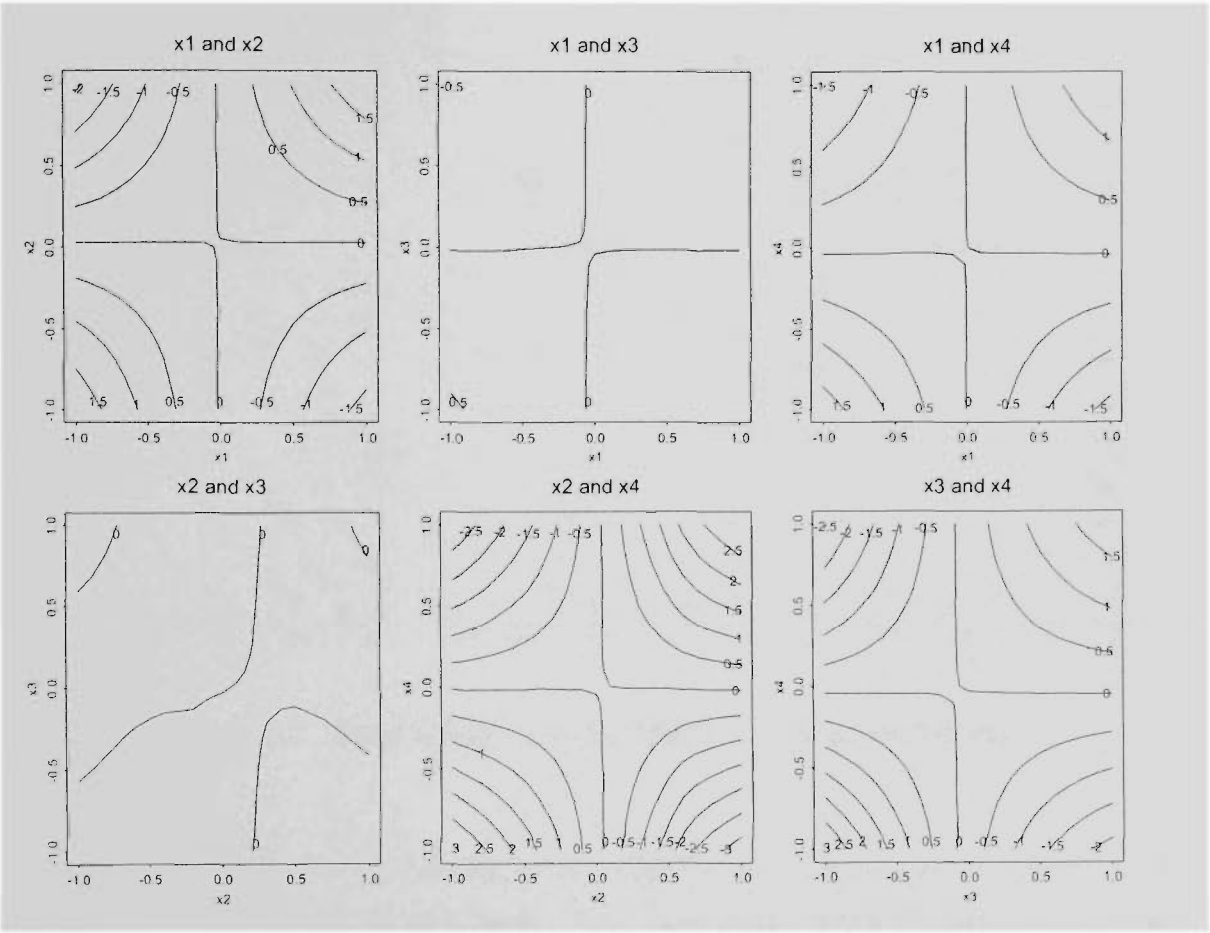

Figure 2.6: Interaction plot for the ASET-B Computer Model.

Another way to determine how important the interaction effects are is to use functional analysis of variance (*functional ANOVA*), an approach applied by Booker (1996) for aeroelastic simulation of the performance of a helicopter rotor. The functional ANOVA, models the overall effect of the covariates as a specified sum of a constant effect, main effects (functions of one covariate) and selected low-order interactions (functions of a few covariates). This decomposition can be used to achieve dimensionality reduction and at the same time retain the flexibility of nonparametric modeling (Huang *et al.* 2000). Use of functional ANOVA

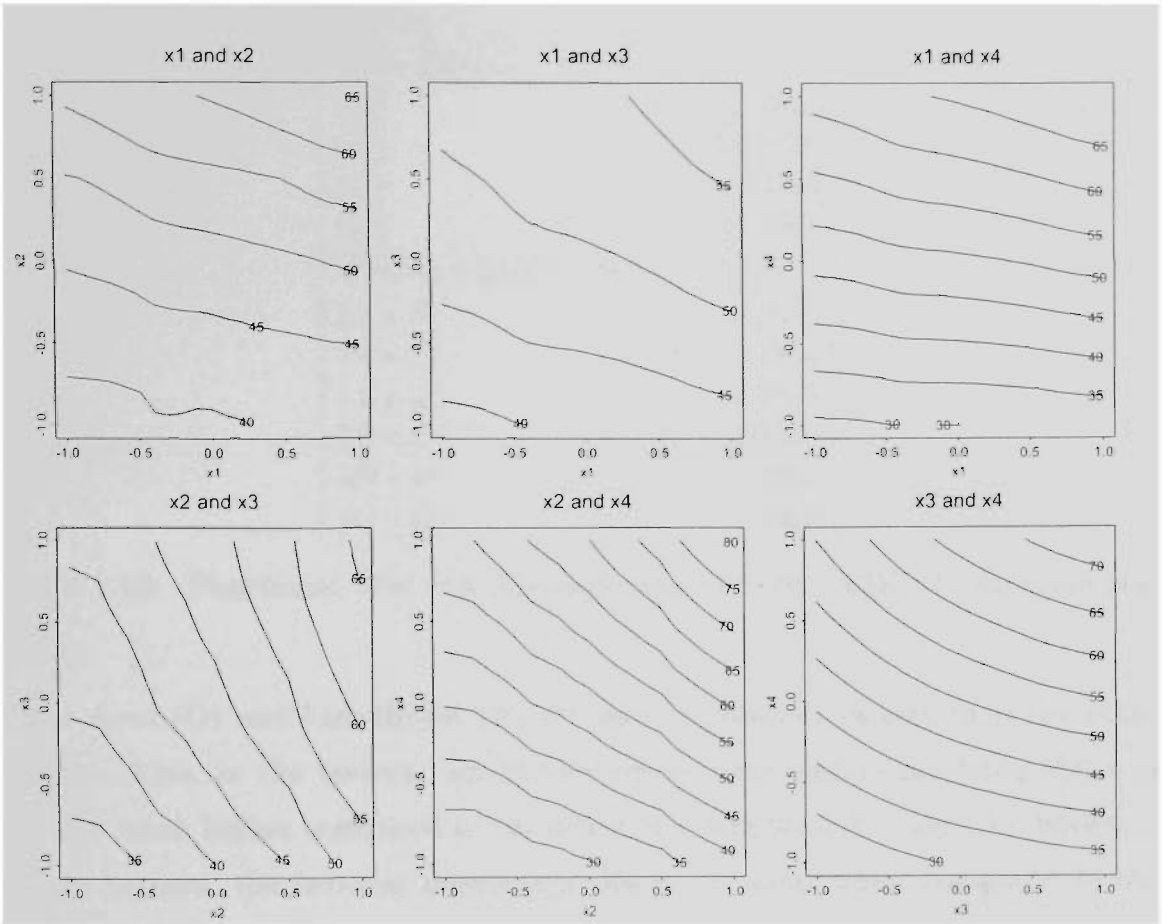

Figure 2.7: Joint Effects for the ASET-B Computer Model.

on sensitivity analysis by Saltelli *et al.* (1999) and Archer *et al.* (1997) is a significant development in this field. The functional ANOVA decomposes the model into contributions of selected input variables. As this decomposition can be applied to any square integral function (Owen, 1992b), this approach was implemented to the function involving estimation of $\mu_i(z_i)$ and $\mu_{i,j}(z_i, z_j)$, which are the estimates for input factors over the design region. The squared variation (integration over the experimental region; say the total amount of squared variation in the model $\hat{Y}$ about its mean is given by $\int \ldots \int (\hat{Y}(x_1, \ldots, x_d) - \hat{\mu})^2 dx_1 \ldots dx_d$, where $\hat{\mu}$ is the function only of the variable x_i. This variation can be decomposed (orthogonally) into the sum of the squared variation from all the main effects and all interactions (two-way and higher) of all the main effects and two-way interactions as indicated in Table 2.10. As presented in Table 2.10, the Room

Source	Estimates
Main Effects	
(A)	4.9003
(B)	76.6982
(C)	30.5498
(D)	215.1293
Two-way Interactions	
$(A * B)$	2.2679
$(A * C)$	0.1406
$(A * D)$	1.5637
$(B * C)$	0.0318
$(B * D)$	6.4065
$(C * D)$	3.5862

Table 2.10: Functional ANOVA decomposition for the ASET-B experimental data.

Floor Area (D) and Fire Height (B) are more significant factors than the other factors. Also, in the two-way interaction effects, the factors involving (D) and (B) are much higher compared to the other two variables, (C) and (A), however, as can be seen, the two-way interaction effects are small when compared to the large main effects.

2.7 Conclusion

ASET-B has been successfully modeled using the methodology introduced by Sacks *et al.* (1989a). The model gives prediction at untried inputs that are very close to the actual responses.

The main effect diagram shows that, over the range studied, the egress time increases almost linearly as the input variables increase. The two-factor interaction diagrams show that non-additivity is quite small for this model. This is supported by the Functional Analysis of Variance.

In later chapters examination of a number of issues such as sample size and augmenting will be undertaken.

Chapter 3

Sample Size Considerations

3.1 Introduction and Background

One of the open problems in computer experiments is how many runs a computer experiment should have. In attempting to answer this question Welch *et al.* (1992), suggested [...] *" taking 10 times the number of active inputs (admittedly a guess)"*. Indeed a method of adding a number of runs to a computer experiment or deciding on a sample size for a computer experiment prior to the experiment would be highly useful for scientists working with computer experiments.

Koehler (1990) performed a study on the behaviour of maximum likelihood estimators (of β, σ^2 and θ) when the sample size is changed. In his study the design space was fixed for different correlation families. He suggested that relatively small sample sizes for computer experiments in terms of maximum likelihood estimators may yield poor results. Further, the study highlighted the side effect of too many samples in which an increase in the sample size will not improve the maximum likelihood estimators. He proposed a rule for the maximum sample size as $n_{max} = 10 * d$, (where d is dimension of input variables of the computer code selected) only for efficiency of estimation and not necessarily for good overall prediction.

This chapter will address the issue of sample size for good prediction by experimenting on four selected computer models. These models are ASET-B (details

are in Chapter 2), Detector Actuation Quasi Steady [**DETACT-QS**] (Evans and Stroup 1985), Circuit Simulation model using the design of a Wheatstone Bridge [**CIRC**] (Box and Fung 1986) and Detector Actuation Time Squared [**DETACT-T2**] (Evans and Stroup 1985) respectively. The details of these computer models will be discussed in the next section.

3.2 Some Details of the Selected Computer Models

3.2.1 DETACT-QS

This computer model which determines the response of fixed temperature and rate of rise heat detectors to fire with energy release rates described by the expression $Q = \alpha t^2$. The complex equations are solved using Newton-Raphson techniques. **DETACT-QS** can be used to predict the actuation time of fixed temperature heat detectors and sprinkler heads subject to a user specified fire. **DETACT-QS** assumes that the thermal device is located in a relatively large area, that only the fire ceiling flow heats the device, and there is no heating from the accumulated hot gases in the room. The required model inputs are the height of the ceiling above the fire, the distance of the detector from the axis (DDA), the actuation temperature of the thermal device (DAT), the response time index (RTI) for the device, and the rate of heat release of the fire. The required range of inputs are listed in Table 3.1. The computer model outputs are the ceiling gas temperature and the device temperature both as a function of time and the time required for device actuation.

DETACT-QS was used as the second four dimensional case with a medium fire growth rate being specified. The computer code output recorded and analysed was the time required for device actuation.

Variable	Variable name	Minimum	Maximum
x_1	Ceiling Height (m)	2.00	4.00
x_2	DDA (m)	1.00	4.00
x_3	DAT (C°)	50.00	75.00
x_4	RTI ($\sqrt{\text{m} * \text{s}}$)	100.00	500.00

Table 3.1: Input Factors for **DETACT-QS** Computer Model

3.2.2 CIRC

Box and Fung (1986), discussed the well known electrical circuit called a Wheatstone Bridge, which is used to determine an unknown resistance ρ_0. This circuit was used by Taguchi(1986) to illustrate the orthogonal array method of parameter design.

A diagramatical illustration of this circuit is shown in Figure 3.1. The components A, C, D and F are fixed resistances; E is a fixed battery voltage; and X is an ammeter reading. If the variable resistance B is adjusted so that there is no flow of current through the ammeter, an estimate y of the unknown resistance ρ_0 can be calculated from the formula

$$y = \frac{BD}{C}.$$

More generally if a current X is flowing through the ammeter, the relation becomes

$$y = \frac{BD}{C} - \frac{X}{C^2 E}\left[A(C+D) + D(B+C)\right]\left[B(C+D) + F(B+C)\right]$$

by which y is related to all the factors A, B, C, D, E, F and X.

In this application of parameter design each of the design resistances A, C, D, and F, and the fixed battery voltage E has an error, corresponding to manufacturing variation, attached to it, and similarly for the battery voltage E, the variable resistance reading B and the ammeter reading X. The aim is to choose the nominal levels of A, C, D, E and F so that the error in determining ρ_0 is minimised.

The method use by Taguchi is to use an inner and an outer array. The inner array involves a 36 run orthogonal array design in the nominal values each at three levels, while the outer array involves a 36 run orthogonal array design in the manufacturing variations, also at three levels. For each run of the inner array the mean, $\overline{y}$, and the standard deviation, s_y, over the outer array points is determined. Taguchi chooses as his criteria the maximisation of

$$SN_T = \left[\frac{\overline{y}^2}{s_y^2} - \frac{1}{36}\right].$$

Box and Fung (1986) use nonlinear programming to equivalently minimise the logarithm of the co-efficient of variation which they show can be estimated more efficiently and directly using numerical differentiation. In the experiment outlined in Table 3.2 and Table 3.3, the Box and Fung (1986) response has been used.

Factor	Factor name	Min	Int	Max
A	Fixed Resistance (ohms)	20.0	100.0	500.0
C	Fixed Resistance (ohms)	2.0	10.0	50.0
D	Fixed Resistance (ohms)	2.0	10.0	50.0
E	Fixed Battery Voltage (Volts)	1.2	6.0	30.0
F	Fixed Resistance (ohms)	2.0	10.0	50.0

Table 3.2: Minimum, Intermediate and Maximum levels of the controllable factors for the **CIRC** Computer Model

Factor	Min	Int	Max
$A(\%)$	-0.3	0	0.3
$B(\%)$	-0.3	0	0.3
$C(\%)$	-0.3	0	0.3
$D(\%)$	-0.3	0	0.3
$E(\%)$	-0.3	0	0.3
$F(\%)$	-0.3	0	0.3
X(mA)	-0.2	0	0.2

Table 3.3: Manufacturing variations of the factors in the **CIRC** computer Model

3.2.3 DETACT-T2

DETACT-T2 is a computer model for calculating the actuation time of thermal

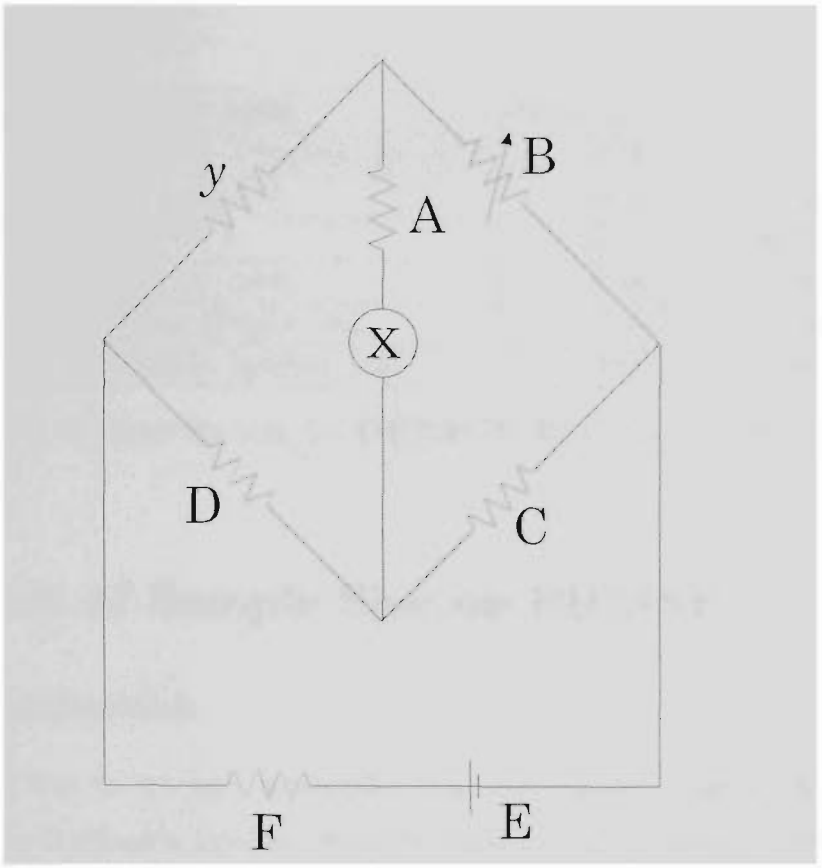

Figure 3.1: Diagram of Wheatstone Bridge.

devices below unconfined ceilings. It can be used to predict the actuation time of fixed temperature and rate of rise heat detectors, and sprinkler heads subject to a user specified fire which grows as the square of time. **DETACT-T2** assumes that the thermal device is located in a relatively large area, that only the fire ceiling flow heats the device and there is no heating from the accumulated hot gases in the room. The required program inputs are the ambient temperature, the response time index (RTI) for the device, detector activation temperature (DAT), the activation and rate of rise (DRR) of the device, height of the ceiling above the fuel, the device spacing and the fire growth rate. The required range of inputs are listed in Table 3.4. The program outputs are the time to device activation and the heat release rate at activation.

The **DETACT-T2** model was used as a six dimensional case and a medium fire growth rate was specified. The output recorded and analysed was the device

activation time.

Variable	Variable name	Minimum	Maximum
x_1	Ambient Temperature (C°)	10.00	25.00
x_2	RTI ($\sqrt{m * s}$)	50.00	500.00
x_3	DAT (C°)	30.00	80.00
x_4	DRR (C/min)	3.00	10.00
x_5	Ceiling Height (m)	2.00	5.00
x_6	Detector Spacing (m)	1.00	5.00

Table 3.4: Input Factors for **DETACT-T2** Computer Model

3.3 Effect of Sample Size on ERMSE

3.3.1 Introduction

In chapter 2 it was shown how a complex computer model could be modelled as the output of a stochastic process using 50 runs of the computer model. Based on these runs predictions for untried inputs of the computer model can be made. The significance of proper guidance for choosing the number of data points for adequate modelling has been discussed by Lucas (1996). Lucas pointed out that for practitioners this would be very useful if they want to comfortably use the design and analysis of computer experiments approach. In this section an experiment on the four computer models described in Chapter 2 and section 3.2 is described which hopefully will shed some light on appropriate sample sizes for a computer experiment.

3.3.2 Methods

Sample sizes of 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 runs were chosen for each computer code. For each sample size three LHDs were generated and the computer models were run at each of the design points. For each LHD the maximum likelihood estimates were calculated using 10 random starting values to improve the chances of converging to the global maximum of the likelihood. Following this the quality of the prediction is determined by calculating the Empirical Root

Mean Square Error (ERMSE) using Equation 2.2 with 1000 randomly selected points over the input region.

3.3.3 Results

As expected, the ERMSE decreases as the sample size increases. Individual results for each computer model are presented in Table 3.5, Table 3.6, Table 3.7 and Table 3.8 respectively. Note that there is significant variability in the quality of the predictions, as measured by the ERMSE, due to the different LHD samples.

	ERMSE			
Sample Size	LHD_1	LHD_2	LHD_3	Mean
10	0.40	0.87	0.74	0.683
20	0.20	0.19	0.19	0.194
30	0.14	0.14	0.22	0.165
40	0.12	0.13	0.11	0.120
50	0.10	0.10	0.08	0.092
60	0.06	0.08	0.07	0.071
70	0.07	0.06	0.07	0.064
80	0.05	0.05	0.08	0.060
90	0.05	0.05	0.04	0.045
100	0.04	0.04	0.03	0.037

Table 3.5: ERMSE for 3 LHDs for Sample Sizes of 10, 20, ..., 100 with the ASET-B Computer Model

3.4 Other Measures of Prediction Quality

The ERMSE is a measure of absolute error. A measure of relative error is

$$\frac{\text{ERMSE}}{\hat{\mu}}$$

where $\hat{\mu}$ is the estimated mean response. Figure 3.2 gives the relationship between this measure and the sample size for the four computer models.

The diagram shows that the four dimensional models have similar relative errors, while as the dimension increases the relative error also increases. While

Sample Size	ERMSE			Mean
	LHD_1	LHD_2	LHD_3	
10	14.65	11.60	12.22	12.820
20	10.49	8.95	9.10	9.513
30	8.49	6.36	7.26	7.368
40	5.05	6.91	5.12	5.690
50	4.19	4.90	4.45	4.513
60	3.69	3.70	3.20	3.530
70	2.97	3.10	3.00	3.026
80	2.08	2.41	2.24	2.241
90	1.98	1.99	1.99	1.980
100	1.71	1.79	1.72	1.708

Table 3.6: ERMSE for 3 LHDs for Sample Sizes of 10.20.....100 with the DETACT-QS Computer Model

the size of the relative error must also depend on the complexity of the model. the general result that the relative error depends partly on the dimension of the model is expected to hold.

Since ERMSE is scale dependent. Lucas (1996) suggested another criterion of prediction quality. R_p^2 . defined as

$$R_p^2 = 1 - \left[\sum [y_i(x) - \hat{y}_i(x)]^2 / \sum [y_i(x) - \overline{y}(x)]^2\right] .$$

Lucas suggested that a good model would require $R_p^2 \geq .99$.

A graph of R_p^2 versus the sample size is given in Figure 3.3. The value of R_p^2 is greater than 0.9 for all computer models only when the sample size is greater than 40.

A paper by Kleijnen and Sargent (2000) on *a methodology for fitting and validating metamodels in simulation* suggests that to include classic DoE and standard measure of fit. such as the R-square coefficient and various cross-validation measures. Further this study includes stagewise DoE and several validation criteria. measures. and estimates. A consideration on this approach is also of a practical use.

Sample Size	ERMSE			Mean
	LHD_1	LHD_2	LHD_3	
10	4.63	4.17	4.80	4.534
20	3.18	2.80	3.74	3.238
30	2.41	2.61	2.56	2.527
40	2.01	2.03	2.04	2.026
50	1.86	1.65	1.90	1.678
60	1.43	1.47	1.59	1.343
70	1.08	1.07	1.09	1.081
80	0.92	0.93	0.93	0.924
90	0.67	0.71	0.73	0.703
100	0.48	0.50	0.49	0.480

Table 3.7: ERMSE for 3 LHDs for Sample Sizes of 10,20,...,100 with the CIRC Computer Model

3.5 Relationship between ERMSE and n

For each computer model we used the experimental data to fit the model

$$\text{ERMSE}^{(\lambda_2)} = a + bn^{(\lambda_1)} + \varepsilon$$

for various values of λ_1 and λ_2. The Box-Cox transformation parameters of n and ERMSE respectively, are given by

$$n^{(\lambda_1)} = \begin{cases} \frac{n^{\lambda_1}-1}{\lambda_1} & \lambda_1 \neq 0 \\ \log(n) & \lambda_1 = 0 \end{cases}$$

$$\text{ERMSE}^{(\lambda_2)} = \begin{cases} \frac{\text{ERMSE}^{\lambda_2}-1}{\lambda_2} & \lambda_2 \neq 0 \\ \log(\text{ERMSE}) & \lambda_2 = 0. \end{cases}$$

The criterion used to measure the quality of the linear fit was R^2. Figure 3.4, Figure 3.5, Figure 3.6, [shaded area in Figure 3.6 depicts various prediction models that are good at ($> 90\%$ for R^2)] and Figure 3.7 show contour graphs of R^2 versus λ_1 and λ_2.

The graphs show that for all four models the best value of $\lambda_1 = 1$, corresponding to no transformation for n. The optimal value of λ_2 differs somewhat for each

Sample Size	ERMSE			Mean
	LHD_1	LHD_2	LHD_3	
10	3.65	2.91	2.45	3.001
20	2.41	2.01	2.11	2.175
30	1.89	1.88	1.81	1.890
40	1.39	1.62	1.47	1.539
50	1.21	1.22	1.22	1.218
60	1.11	1.03	1.08	1.098
70	0.99	1.00	0.99	0.999
80	0.86	0.90	0.80	0.855
90	0.79	0.80	0.70	0.763
100	0.61	0.71	0.61	0.644

Table 3.8: ERMSE for 3 LHDs for Sample Sizes of 10,20,....,100 with the DETACT-T2 Computer Model

of the models. However a compromise value of $\lambda = -0.25$ appears to work well and hence a model relating the sample size to ERMSE is given by

$$\text{ERMSE}^{-0.25} = a + bn. \tag{3.1}$$

Figure 3.8 shows a plot of $\text{ERMSE}^{-0.25}$ versus the sample size for all four computer models. The graph shows that the linear relationship fits well, perhaps less well for ASET-B and CIRC. It was not clear as to whether the behaviour of the fitted model to the ASET-B and CIRC computer models are a contribution of sample sizes, design space selected and/or the dimension of the ASET-B and CIRC computer model itself. Another observation of interest is that with this transformation the variability due to different LHDs is approximately constant independent of the sample size.

The intercepts, slopes and variabilities of the fitted relationships are given in Table 3.9. The Table shows that the slopes are clearly different for all the models.

3.6 Discussion

In this chapter an experiment on four computer models was conducted. In the experiment each computer experiment was modelled using Latin Hypercubes of

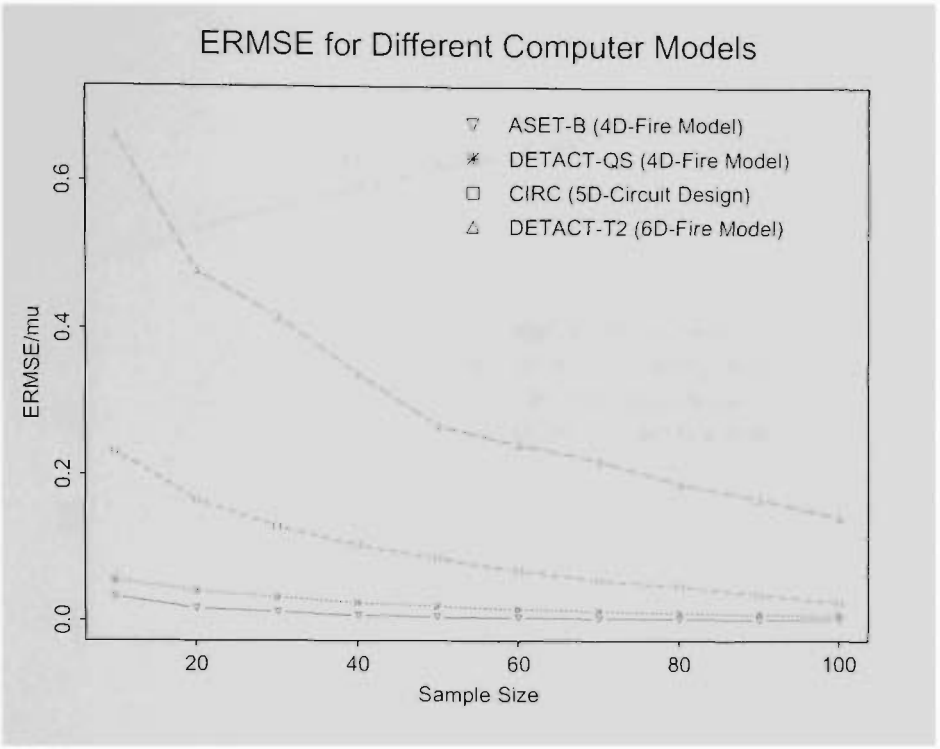

Figure 3.2: Scaled ERMSE versus sample size for the four computer models.

various sizes from 10 to 100 runs. The prediction quality was measured using the ERMSE.

Analysis of the data showed that by transforming the ERMSE response, linear relationships could be found relating the transformed response to the sample size. The best transformed response was found to be $\text{ERMSE}^{-0.25}$.

Although good fits have been obtained for all models, the fits show that the slope parameter is different for all models ranging three-fold. Similarly the intercept parameter is different for all models. If a guess for the slope and intercept

Computer Code	a	b	σ
ASET-B	1.1997	0.0112	0.01988456
DETACT-QS	0.4930	0.0039	0.01449088
CIRC	0.6254	0.0052	0.02554453
DETACT-T2	0.7462	0.0037	0.01957868

Table 3.9: Estimates of a, b and σ
in Model (Equation 3.1) for the four computer models.

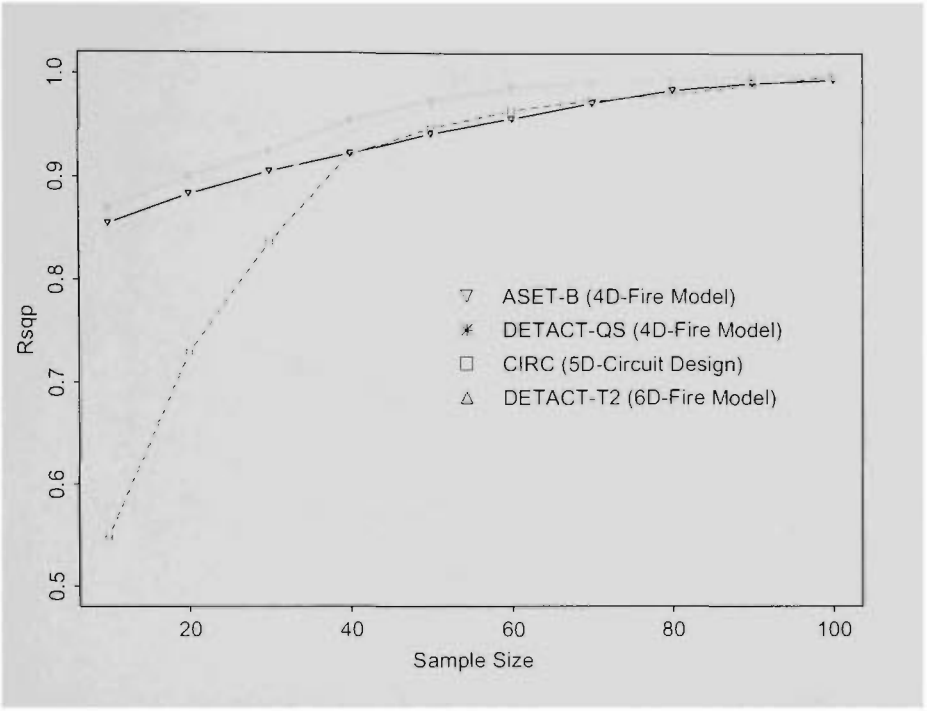

Figure 3.3: R_p^2 versus sample size for the four computer models.

could be given then a formula for the sample size for a desired ERMSE can be given as

$$N = \frac{\text{ERMSE}^{-0.25} - a}{b}.$$

For example if a was expected to be 0.8 and b to be 0.01, the sample size for a ERMSE of 0.025 would be about 74.

The obvious difficulty is that, since the slopes and intercepts differ for each model, it will be difficult for an experimenter to guess the likely values of the parameters a and b. However, further experimentation along the lines of that done in this chapter may give an experimenter an indication of the range of values that are usual for similar types of computer experiments as the one she or he is using.

The results also show that, as far as prediction is concerned, taking ten times the number of active factors may or may not work well depending on the computer model. With the R_p^2 criterion all four computer models only achieved $R_p^2 \geq 0.9$ when $N \geq 40$. To achieve $R_p^2 \geq 0.99$, as Lucas suggested, would require much

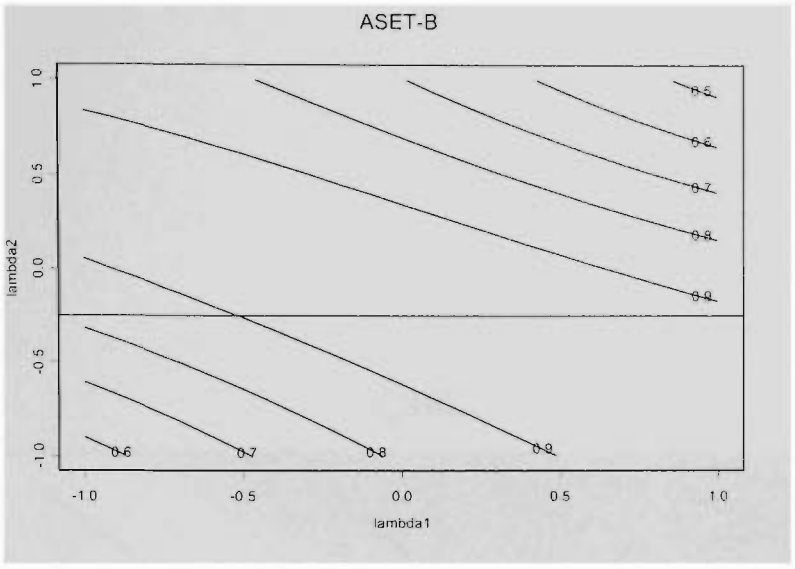

Figure 3.4: Contour Plot of R^2 versus λ_1 and λ_2 for ASET-B.

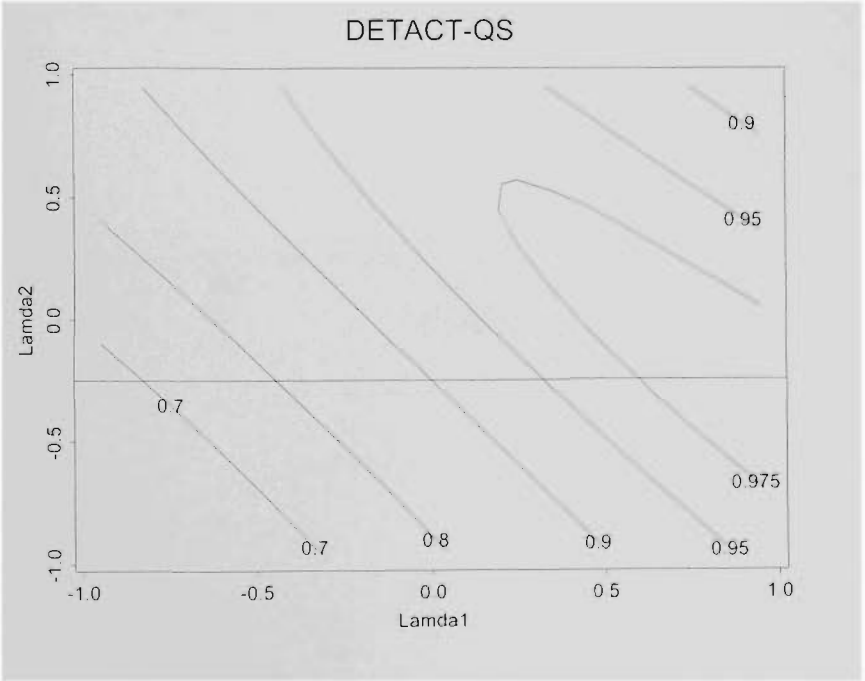

Figure 3.5: Contour Plot of R^2 versus λ_1 and λ_2 for DETACT-QS.

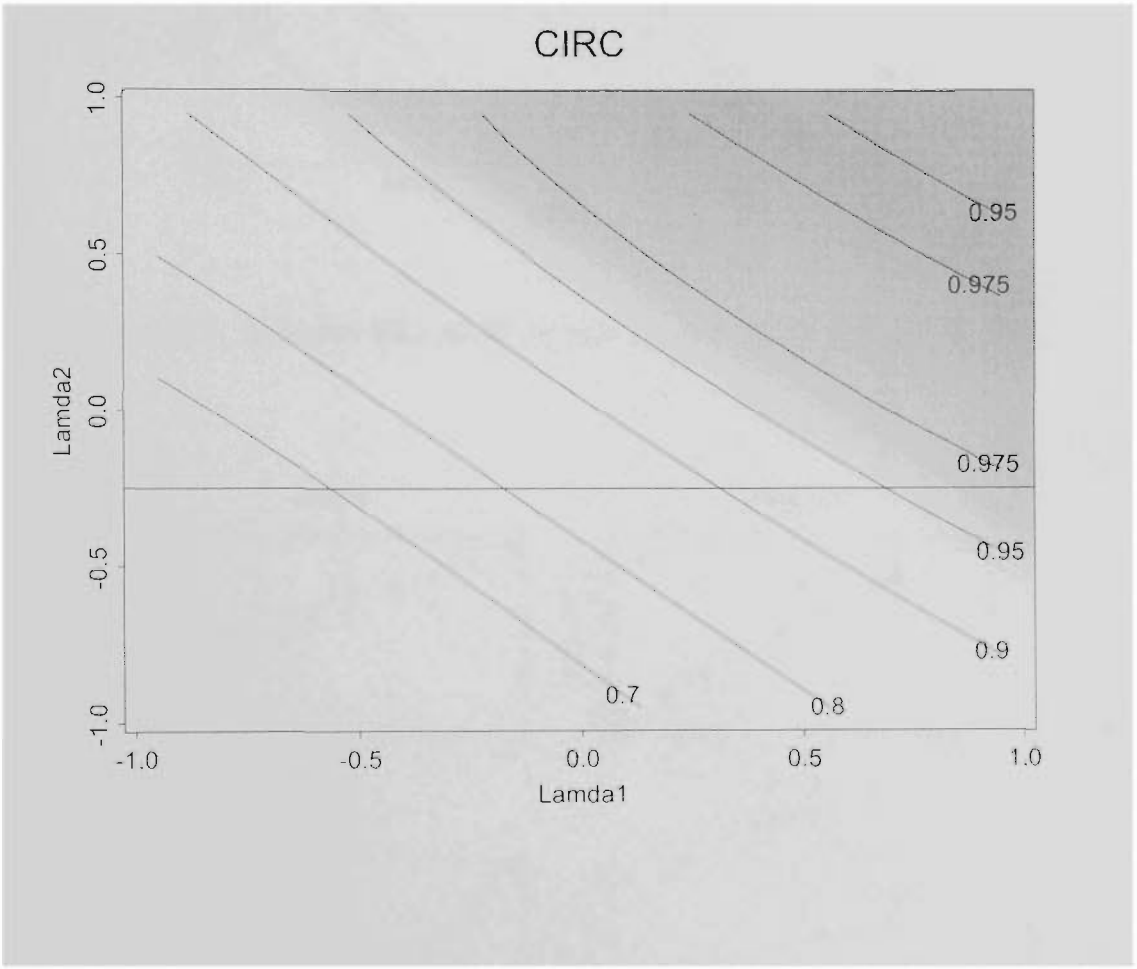

Figure 3.6: Contour Plot of R^2 versus λ_1 and λ_2 for CIRC.

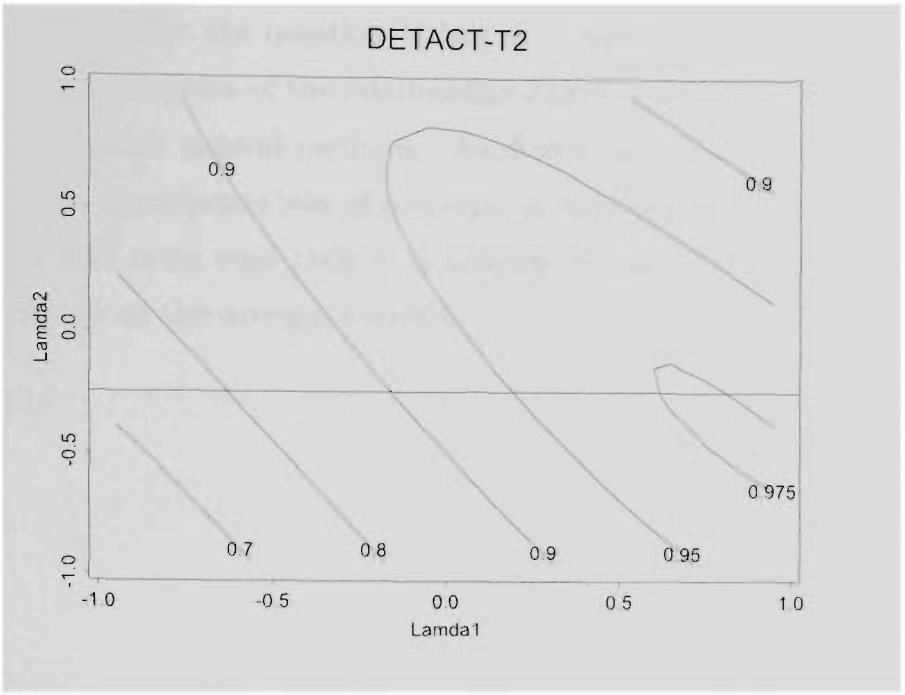

Figure 3.7: Contour Plot of R^2 versus λ_1 and λ_2 for DETACT-T2.

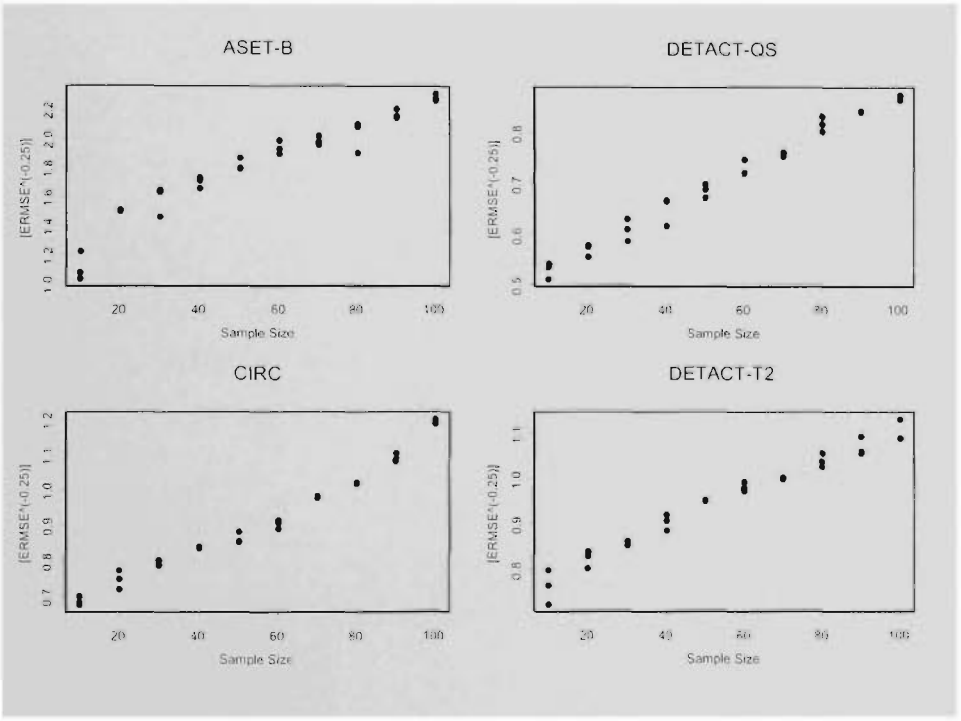

Figure 3.8: Plots of Linear Relationship for Different Computer Models.

greater sample sizes.

In the next chapter the question of how to augment an experiment will be discussed. One application of the relationship found in this chapter will be given as well as some more general methods. As shown in this chapter it is difficult to decide on the appropriate size of a computer experiment in advance but if it is possible to add extra runs then it is possible to control the precision of the predictions based on the surrogate model.

Chapter 4

Augmenting Computing Experiments

4.1 Introduction

In the previous chapter some considerations as to the size of a computer experiment were applied. After the completion of the experiment, further runs may need to be added. In this chapter, investigations will be focused on how to augment a computer experiment.

Specifically two situations will be studied:

1. *When one run is added to a computer experiment.* This would be helpful if one point is needed to add into a particular case when possibly re-running the computer code is highly expensive or when it is possible to add points individually and review the results point by point.

2. *When more than one run is added to a computer experiment.* This might apply when there is no opportunity of adding points sequentially and reviewing the results. Clearly the sequential approach has many advantages if it is practically feasible.

In addition, a method based on the results of the previous chapter is also suggested.

4.2 Adding one run to a computer experiment

After a computer experiment, we can generate both a prediction for each point over the range of interest as well as the MSE of the prediction which describes how uncertain we are in our prediction. The MSE is zero for the points already contained in the computer experiments. A good method for adding a single run would be to calculate the MSE at each of a potential set of points and choose the point with the maximum MSE. This method is similar to the simple algorithm developed by Dykstra (1971). The candidate set of points can be a Latin Hypercube, a grid, any other Orthogonal Array or Computer Design or even a Random Set of points over the experimental region. For illustration purposes a 20 run LHD is used as the initial design and a 10 run LHD as the candidate set although in practice the candidate set should be much larger than the initial design.

Table 4.1 gives the coded design points of the initial 20 run LHD with corresponding egress times from ASET-B.

Modelling the output as a realization of a stochastic process gave the results given in Table 4.2.

Table 4.3 gives the coded candidate points from a 10 run ASET-B again with the corresponding egress times. A pairs plot of the original design points and the candidate points is given in Figure 4.1.

For each of the candidate points the MSE was calculated using Equation 2.3. Note, instead of using the Equation 2.3, by ignoring the random character of the estimated correlation matrix R and the resulting *Kriging* weights. Kleijnen and Van Beers (2003) use cross-validation to estimate the correct variance.

The MSEs are given in the column labelled "1st" in Table 4.4. The candidate point with the largest MSE is point 6. If this point is added and the model refitted the MSEs at the remaining points are given in the column labelled "2nd" in Table 4.4. With the result of the additional point 6 added the point with the largest MSE is now Point 9.

Sometimes it is possible to proceed sequentially. The remaining columns of Table 4.4 give the results when the process above is repeated another three times.

$19x_1$	$19x_2$	$19x_3$	$19x_4$	Egress Time
−5	−1	13	−7	45.00
−9	11	11	−3	55.00
−17	1	−9	−11	34.00
19	9	7	−19	36.25
7	−9	19	11	61.25
−19	−11	−19	11	38.33
5	−3	1	13	60.00
−7	3	−11	19	60.00
11	13	−5	−5	52.50
−3	17	−7	17	72.50
17	19	−17	−9	48.33
−11	15	−1	−15	39.00
1	7	−13	5	52.50
15	−7	15	−15	39.00
1	7	−13	5	52.50
15	−7	15	15	66.67
−1	−19	−15	9	40.00
13	−15	−3	−13	30.00
−15	5	17	1	55.00
9	−17	3	3	45.00
3	−13	9	−1	45.00
−13	−5	5	−17	30.00

Table 4.1: Coded design points of the Initial Design with Egress Times from ASET-B.

Hence after adding point 9 and refitting the model. Point 1 gives the largest MSE. and once this point is added the next point to add is Point 2 followed by Point 8.

In order to determine how far to continue the process the average MSE of the candidate points is given in the bottom row of the table. The average includes the augmenting points where the MSE must be zero. When the average reaches a suitably small value then no further points are added.

i	$\hat{p}_i$	$\hat{\theta}_i$
1	2	.0040
2	1.97	.0128
3	1.996	.0052
4	1.998	.0450

$$\begin{aligned} \hat{\beta} &= 42.7608 \\ \hat{\sigma}^2 &= 2820.2617 \end{aligned}$$

Table 4.2: Results of Modelling ASET-B using the 20 run design given in Table 4.1

$9x_1$	$9x_2$	$9x_3$	$9x_4$	Egress Time
−1	−7	1	9	56.67
9	3	−7	3	55.00
−9	1	−3	1	46.67
−7	−1	9	−1	49.00
5	−3	5	−5	39.00
7	9	7	5	85.00
3	5	3	−3	51.67
1	7	−5	−9	32.50
−5	−5	−9	−7	24.17
−3	−9	−1	−7	50.00

Table 4.3: Coded potential augmenting runs with Egress Times from ASET-B.

4.3 Adding more than one run to a computer experiment

When more than one augmenting point is required an alternative method would be to generate the variance-covariance matrix of the potential points and choose that subset which has the largest determinant. For this purpose an expression for the covariance of two prediction errors needs to be generated. The following proposition is used:

Proposition 4.3.1 *For potential points selected as* $\mathbf{u}$ *and* $\mathbf{v}$ *the prediction variance-covariance matrix can be generated by*

$$Cov(\hat{y}(\mathbf{u}), \hat{y}(\mathbf{v})) = \sigma^2 \left[r_{\mathbf{uv}} - \begin{pmatrix} 1 & \mathbf{r}^T(\mathbf{u}) \end{pmatrix} \begin{pmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{R}_D \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mathbf{r}(\mathbf{v}) \end{pmatrix} \right]$$

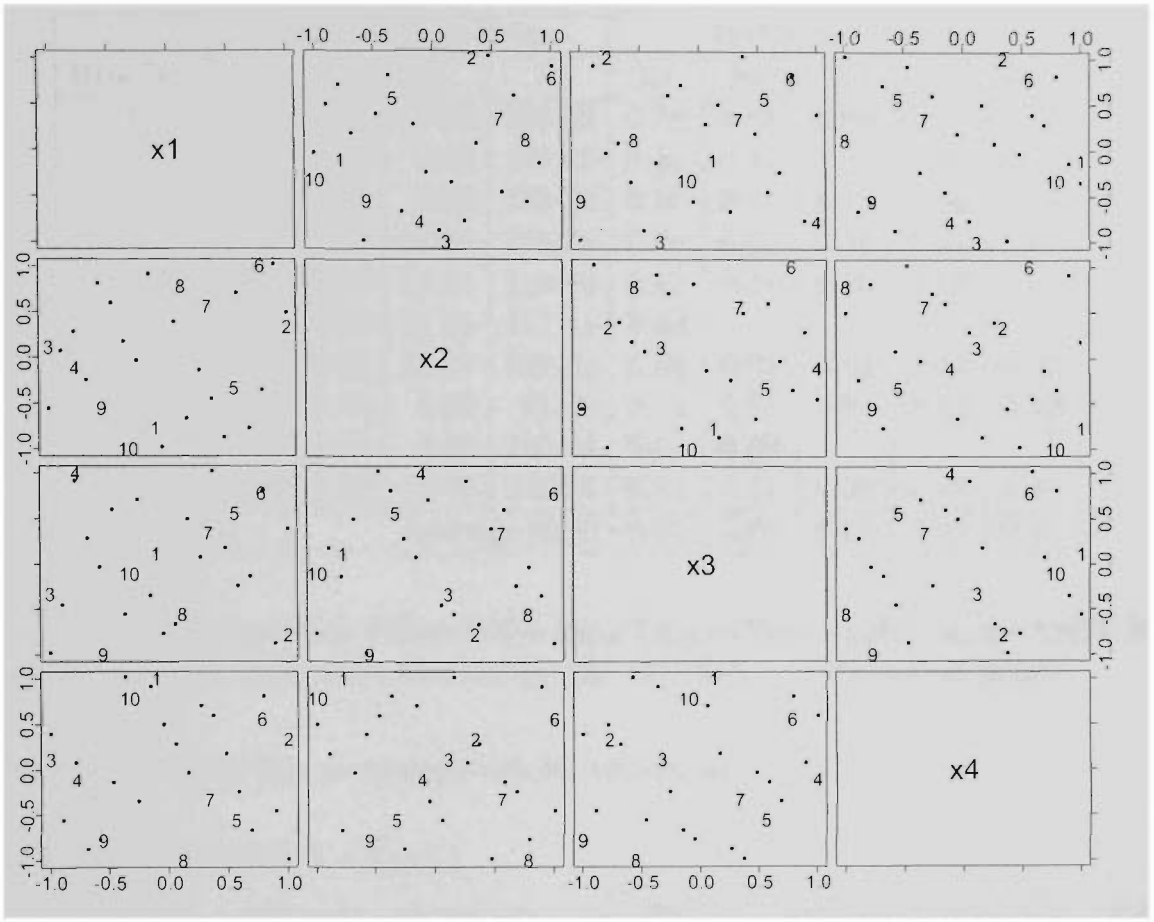

Figure 4.1: Original 20 points from a LHD plus 10 labeled candidate points from a 10 run LHD

where the correlation of the zs at $\mathbf{u}$ *and* $\mathbf{v}$ *is* $\mathbf{r_{uv}}$. *the correlation between the zs at the design points and at* $\mathbf{u}$ *is*

$$\mathbf{r}(\mathbf{u}) = [R(\mathbf{x}_1, \mathbf{u}), \ldots . R(\mathbf{x}_n, \mathbf{u})]^T$$

and the correlation between the zs at the design points and at $\mathbf{v}$ *is*

$$\mathbf{r}(\mathbf{v}) = [R(\mathbf{x}_1. \mathbf{v}), \ldots , R(\mathbf{x}_n. \mathbf{v})]^T.$$

	Ten Candidate Points				MSE$[\hat{y}(x)]/sigma^2$				
Run No.	x_1	x_2	x_3	x_4	1st	2nd	3rd	4th	5th
1	0.73	1.22	10.22	256.00	0.78	0.67	**0.68**		
2	0.90	2.33	8.44	197.67	0.44	0.51	0.50	**0.36**	
3	0.60	2.11	9.33	178.22	0.16	0.12	0.07	0.06	0.04
4	0.63	1.89	12.00	158.78	0.06	0.05	0.04	0.03	0.02
5	0.83	1.67	11.11	119.89	0.17	0.11	0.10	0.08	0.05
6	0.87	3.00	11.56	217.11	**1.54**				
7	0.80	2.56	10.67	139.33	0.10	0.05	0.04	0.04	0.02
8	0.77	2.78	8.89	81.00	0.54	0.50	0.38	0.33	**0.23**
9	0.67	1.44	8.00	100.44	0.68	**0.94**			
10	0.70	1.00	9.78	236.56	0.72	0.61	0.58	0.12	0.10
				Average MSE	0.52	0.36	0.24	0.10	0.05

Table 4.4: Ten Candidate Points following a Twenty Points LHD for the ASET-B Fire Model: and MSE of Prediction for the 1st, 2nd, ..., 5th added points.

Proof: The prediction covariance can be written as

$$
\begin{aligned}
& E\{[Y(\mathbf{u}) - \hat{Y}(\mathbf{u})][Y(\mathbf{v}) - \hat{Y}(\mathbf{v})]\} \\
&= E\{[Y(\mathbf{u}) - E(Y(\mathbf{u})) + E(\hat{Y}(\mathbf{u})) - \hat{Y}(\mathbf{u})][Y(\mathbf{v}) - E(Y(\mathbf{v})) + E(\hat{Y}(\mathbf{v})) - \hat{Y}(\mathbf{v})]\} \\
&= \mathcal{C}(Y(\mathbf{u}), Y(\mathbf{v})) - \mathcal{C}(Y(\mathbf{u}), \hat{Y}(\mathbf{v})) - \mathcal{C}(Y(\mathbf{u}), \hat{Y}(\mathbf{v}) + \mathcal{C}(\hat{Y}(\mathbf{u}), \hat{Y}(\mathbf{v})) \\
&= \sigma^2 r_{\mathbf{u}\mathbf{v}} - \mathcal{C}(Y(\mathbf{u}), \mathbf{BY})) - \mathcal{C}(Y(\mathbf{v}), \mathbf{AY}) + \mathcal{C}(\mathbf{AY}, \mathbf{BY}) \\
&= \sigma^2(r_{\mathbf{uv}} - \mathbf{r}^T(\mathbf{u})\mathbf{B}^T - \mathbf{r}^T(\mathbf{v})\mathbf{A}^T + \mathbf{A}\mathbf{R}_D\mathbf{B}^T)
\end{aligned}
$$

where

$$
\mathbf{A} = (1^T\mathbf{R}_D^{-1}1)^{-1}1^T\mathbf{R}_D^{-1} + \mathbf{r}(\mathbf{u})^T\mathbf{R}_D^{-1} - \mathbf{r}(\mathbf{u})^T\mathbf{R}_D^{-1}1(1^T\mathbf{R}_D^{-1}1)^{-1}1^T\mathbf{R}_D^{-1}
$$

$$
\mathbf{B} = (1^T\mathbf{R}_D^{-1}1)^{-1}1^T\mathbf{R}_D^{-1} + \mathbf{r}(\mathbf{v})^T\mathbf{R}_D^{-1} - \mathbf{r}(\mathbf{v})^T\mathbf{R}_D^{-1}1(1^T\mathbf{R}_D^{-1}1)^{-1}1^T\mathbf{R}_D^{-1}
$$

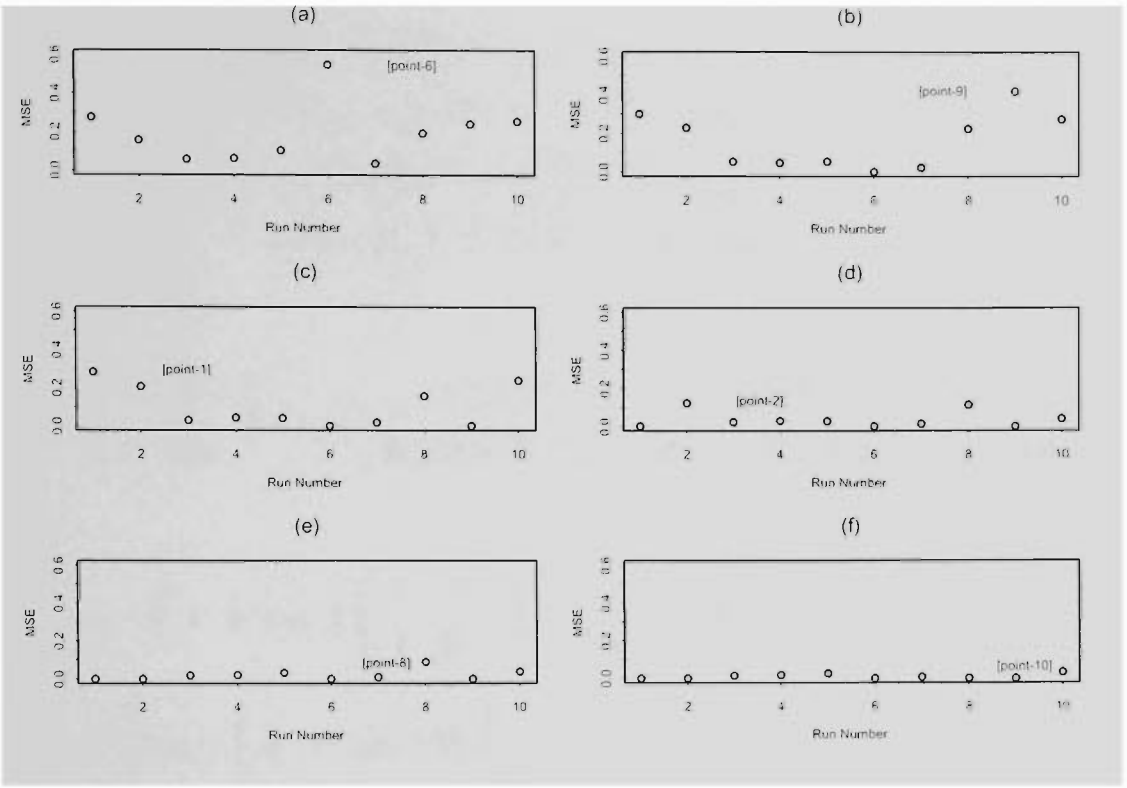

Figure 4.2: Mean Square Error of Prediction adding one point at a time.

Now,

$$
\begin{aligned}
\mathbf{r}^T(\mathbf{u})\mathbf{B}^T &= \mathbf{r}^T(\mathbf{u})\mathbf{R}_D^{-1}\mathbf{1}(\mathbf{1}^T\mathbf{R}_D^{-1}\mathbf{1})^{-1} + \mathbf{r}^T(\mathbf{u})\mathbf{R}_D^{-1}\mathbf{r}(\mathbf{v}) \\
&\quad - \mathbf{r}^T(\mathbf{u})\mathbf{R}_D^{-1}\mathbf{1}(\mathbf{1}^T\mathbf{R}_D^{-1}\mathbf{1})^{-1}\mathbf{1}^T\mathbf{R}_D^{-1}\mathbf{r}(\mathbf{v}) \\
\mathbf{r}^T(\mathbf{v})\mathbf{A}^T &= \mathbf{r}^T(\mathbf{v})\mathbf{R}_D^{-1}\mathbf{1}(\mathbf{1}^T\mathbf{R}_D^{-1}\mathbf{1})^{-1} + \mathbf{r}^T(\mathbf{v})\mathbf{R}_D^{-1}\mathbf{r}(\mathbf{u}) \\
&\quad - \mathbf{r}^T(\mathbf{v})\mathbf{R}_D^{-1}\mathbf{1}(\mathbf{1}^T\mathbf{R}_D^{-1}\mathbf{1})^{-1}\mathbf{1}^T\mathbf{R}_D^{-1}\mathbf{r}(\mathbf{u}) \\
\mathbf{A}\mathbf{R}_D\mathbf{B}^T &= (\mathbf{1}^T\mathbf{R}_D^{-1}\mathbf{1})^{-1} + \mathbf{r}^T(\mathbf{u})\mathbf{R}_D^{-1}\mathbf{r}(\mathbf{v}) \\
&\quad - \mathbf{r}^T(\mathbf{u})\mathbf{R}_D^{-1}\mathbf{1}(\mathbf{1}^T\mathbf{R}_D^{-1}\mathbf{1})^{-1}\mathbf{1}^T\mathbf{R}_D^{-1}\mathbf{r}(\mathbf{v})
\end{aligned}
$$

Hence,

$$\begin{aligned} & E\{[Y(\mathbf{u}) - \hat{Y}(\mathbf{u})][Y(\mathbf{v}) - \hat{Y}(\mathbf{v})]\} \\ &= \sigma^2 \left[r_{\mathbf{uv}} + (1^T \mathbf{R}_D^{-1} 1)^{-1} - \mathbf{r}^T(\mathbf{u}) \mathbf{R}_D^{-1} 1 (1^T \mathbf{R}_D^{-1} 1)^{-1} \right. \\ &\quad + \mathbf{r}^T(\mathbf{v}) \mathbf{R}_D^{-1} 1 (1^T \mathbf{R}_D^{-1} 1)^{-1} 1^T \mathbf{R}_D^{-1} \mathbf{r}(\mathbf{u}) \\ &\quad \left. - \mathbf{r}^T(\mathbf{v}) \mathbf{R}_D^{-1} 1 (1^T \mathbf{R}_D^{-1} 1)^{-1} - \mathbf{r}^T(\mathbf{v}) \mathbf{R}_D^{-1} \mathbf{r}(\mathbf{u}) \right]. \end{aligned}$$

Since

$$\mathbf{W} = \begin{pmatrix} 0 & 1^T \\ 1 & \mathbf{R}_D \end{pmatrix}^{-1} = \begin{pmatrix} -(1^T \mathbf{R}_D^{-1} 1)^{-1} & (1^T \mathbf{R}_D^{-1} 1)^{-1} 1^T \mathbf{R}_D^{-1} \\ \mathbf{R}_D^{-1} 1 (1^T \mathbf{R}_D^{-1} 1)^{-1} & \mathbf{R}_D^{-1} - \mathbf{R}_D^{-1} 1 (1^T \mathbf{R}_D^{-1} 1)^{-1} 1 \mathbf{R}_D^{-1} \end{pmatrix}$$

then

$$\begin{aligned} & \sigma^2 \left[r_{\mathbf{uv}} - \begin{pmatrix} 1 & \mathbf{r}^T(\mathbf{u}) \end{pmatrix} \begin{pmatrix} 0 & 1^T \\ 1 & \mathbf{R}_D \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mathbf{r}(\mathbf{v}) \end{pmatrix} \right] \\ &= \sigma^2 \left[r_{\mathbf{uv}} - \begin{pmatrix} 1 & \mathbf{r}^T(\mathbf{u}) \end{pmatrix} \mathbf{W} \begin{pmatrix} 1 \\ \mathbf{r}(\mathbf{v}) \end{pmatrix} \right] \\ &= \sigma^2 \left[r_{\mathbf{uv}} - (1^T \mathbf{R}_D^{-1} 1)^{-1} + \mathbf{r}(\mathbf{u}) \mathbf{R}_D^{-1} 1 (1^T \mathbf{R}_D^{-1} 1)^{-1} + (1^T \mathbf{R}_D^{-1} 1)^{-1} 1^T \mathbf{R}_D^{-1} \mathbf{r}(\mathbf{v}) \right. \\ &\quad \left. + \mathbf{r}^T(\mathbf{u}) \mathbf{R}_D^{-1} \mathbf{r}^T(\mathbf{v}) - 1^T \mathbf{R}_D^{-1} 1 (1^T \mathbf{R}_D^{-1} 1)^{-1} 1^T \mathbf{R}_D^{-1} \mathbf{r}(\mathbf{v}) \right] \end{aligned}$$

hence proving the result.

Corollary 4.3.1 *The prediction variance at a site* $\mathbf{x}$.

$$E\{[Y(\mathbf{x}) - \hat{Y}(\mathbf{x})]^2\}$$

is given by

$$\sigma^2 \left[1 - \begin{pmatrix} 1 & \mathbf{r}^T(\mathbf{x}) \end{pmatrix} \begin{pmatrix} 0 & 1^T \\ 1 & \mathbf{R}_D \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mathbf{r}(\mathbf{x}) \end{pmatrix} \right].$$

Using this method a number of points could be added at a time. Figure 4.3 gives the determinant of the variance-covariance matrix of the prediction errors for all sets of five runs out of the ten candidate points from the LHD. In this case the best set of five points is also the set $\{1, 2, 6, 8, 9\}$. Clearly there will be other cases where the two points are fitted sequentially. When points are fitted one at a time the first five points are $\{6, 9, 1, 2, 8\}$ (see Figure 4.2).

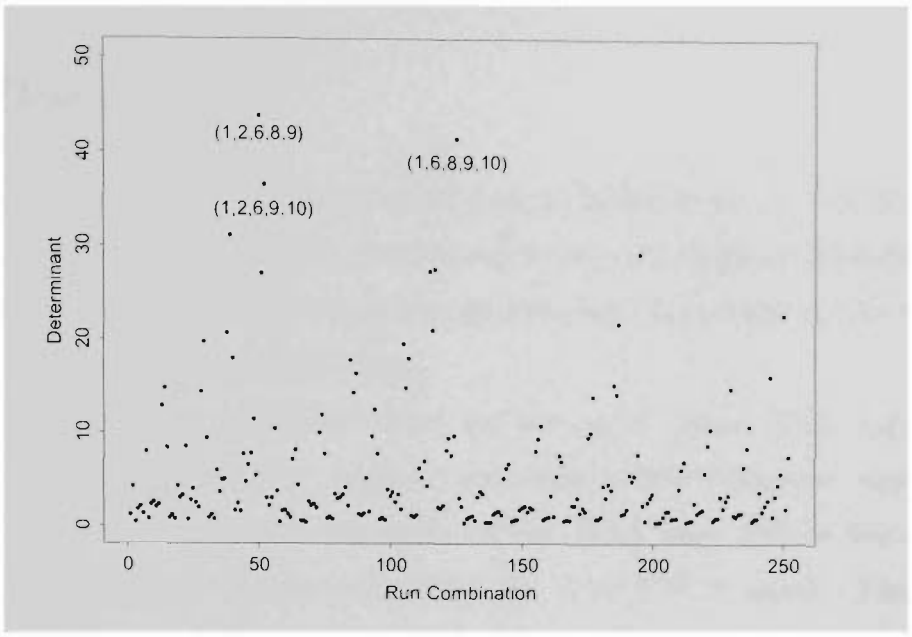

Figure 4.3: Determinant of the variance-covariance matrix of predictions for all possible run combinations of five runs from the candidate set of 10 runs.

4.4 Another Possible Approach

Equation 3.1 gives an approximate straight line relationship between the $\text{ERMSE}^{-0.25}$ and the sample size. The intercept, a, and the slope, b, in the equation depend on the particular computer model.

The parameters a and b are unknown before the experiment. Assume that a $2N + 1$ run computer experiment is to be conducted. Rather than using a $2N + 1$ run Latin Hypercube design it might be worthwhile using an N run Latin Hypercube and an $N + 1$ run Latin Hypercube. Three fits are produced: one based on the N run Latin Hypercube, one based on the $N + 1$ run Latin Hypercube and one based on the combined design and for each fit the ERMSE is calculated.

Using the three ERMSE values, linear regression could be used to estimate a and b, and the sample size sufficient for a particular level of the ERMSE could be determined. This method ignores the dependence between the ERMSE values and the fact that neither the combined design nor augmented design are Latin

Hypercube Designs, but still may give a useful guideline to experimenters.

4.5 Discussion

One of the advantages of the method advocated by Sacks et. al. (1989b) is that the predictions of the response at untried inputs are accompanied by estimates of the size of the prediction error expected. In this way experimenters are informed as to the usefulness of the predictions.

In cases where the prediction errors are too great, there is the opportunity to improve the predictions by adding extra runs of the computer experiment. By using such runs the prediction error at the extra runs will be become zero and in a neighbourhood of the extra runs the error will be small. These extra runs will also have an impact on the predictions and their errors elsewhere in the experimental region.

The question of where to add the extra runs is one of practical importance. Sometimes running the code is just as expensive and time consuming as carrying out a physical experiment. A number of ways of deciding on the location of the extra points can be suggested. Two such methods have been examined in this chapter.

The first method is useful when one point at a time can be added, the results evaluated and then the decision of where the next point can be made. The method takes as the augmenting point that point which gives the maximum prediction error. This method is easily explained to experimenters since points are added where the predictions are worst in terms of size of the prediction error. The procedure can be continued until the average or maximum prediction error is sufficiently small.

If more than one point can be added but there is no opportunity to evaluate the computer model after each point then a modification of the first method can be used. This modification takes the prediction at the point with maximum error as if it was an actual result. There is no need to update the maximum likelihood estimates but the prediction errors should be updated, leading to another point

being selected as the next point to be added as it has the maximum mean square error of prediction. Again this procedure can be continued until the maximum mean square error of prediction is sufficiently small.

Adding points one by one may lead to suboptimal results. An alternative method analogous to those often used for physical experiments is to use the determinant of the prediction variance-covariance matrix allowing the best choice of groups of points. Just as the prediction variance can be calculated so too can the prediction covariance using a generalisation of the well known formula for the mean square error of prediction.

A method based on the result of the previous chapter has also been suggested. This method relies on using two separate designs and gives a prediction of the sample size for a particular desired level of the ERMSE.

Alternatively, it is worth noting that some development aspects of Bayesain Global Optimization method discussed by Schonlau (1997), where to determine how global versus local the search will be when additional parameters are introduced. Given the correlation parameters, the expected improvement criterion optimally choose where to sample one point according to an average case analysis. The paradigm of the average case analysis, given the correlation parameters, thus ultimately determines the balance between the global and local components of the search. When the correlation parameters are poorly estimated, an average case analysis is not sensible, and typically the search is too local. "For many applications sampling one point at a time is unrealistic. For one, unless the sampling can be computer automated it is very time consuming. Second, it may also be more cost effective to have only a few stages where at each stage a number of points are sampled. In other words, sampling m point at a time may be a lot cheaper than sampling at m stage (Schonlau *et al.*, 1998).

An another approach discussed by Jones *et al.* (1998), based on fitting response surfaces to data collected by evaluating the objective and constraint functions at a few points is of a significant development in augmenting computer experiments. These response surfaces are then used to visualize input output relationships, estimate the location of the optimum, and suggest points where

additional function evaluations may help improve this estimate. This approach not only provides an estimate of the optimal point, but also facilitates the development of intuition and understanding about what is going on in the model (Jones *et al.*, 1998).

Finally, a method developed on the *leave-one-out* approach employed by Kleijnen and van Beers (2003) to estimate the true I/O function through $cross-validation$ is worth noting. In this approach, the authors obtained estimation of the uncertainty of output at input combinations not yet observed, by successively deleting one of the I/O observations already simulated. For more information on cross-validation see Stone (1974) and for an update see Meckesheimer *et al.*, (2002) and Mertens (2001).

Chapter 5

A Model for Computer Models

5.1 Introduction

Computer Experiments, consisting of a number of runs of a computer model with different inputs, are now commonplace in scientific research. Based on running the computer model at a set of conditions, a surrogate model can be built to give a good approximation to the output of the computer model at untried inputs. The quality of the prediction depends on the sample size of the experiment and the complexity of the computer model.

This chapter describes an experiment on this process. The inputs of the computer experiment are the parameters that describe the complexity of the computer model and, the sample size and the output is the empirical mean square error of the prediction of the fitted surrogate computer model. The results show how the sample size of the experiment and complexity of the computer model impact on the prediction quality of the surrogate computer model.

5.2 Simulating ASET-B

In Chapter 2 the ASET-B computer model was introduced. A 50 run computer experiment was conducted and the parameters of the surrogate computer model, using the methodology of Sacks *et al.* (1989b), were estimated.

n	σ^2	ln(EMSE)				
10	1	−5.435	−7.126	−6.932	−5.764	−6.460
	123.556	−0.810	−0.275	−1.060	−1.070	−1.099
30	1	−8.800	−8.912	−8.469	−8.910	−8.467
	123.556	−2.846	−4.279	−4.030	−3.442	−3.591
50	1	−8.989	−9.436	−9.763	−8.944	−9.768
	123.556	−4.351	−5.294	−4.750	−4.581	−4.667
70	1	−10.518	−9.926	−10.748	−10.591	−10.749
	123.556	−5.261	−5.604	−5.886	−5.590	−5.511

Table 5.1: Results for experiment on surrogate ASET-B model.

In Chapter 3 an experiment on the ASET-B computer model, as well as three other computer models, was conducted. In the experiment the sample size was varied from 10 to 100 runs and the Empirical Root Mean Square Error of the fitted surrogate computer models was found.

In this section a similar experiment to that of chapter 3 will be conducted using the surrogate computer model. Two sets of simulations will be conducted. In the first set the θ and p parameters found in Table 2.9 will be used to estimate the variance-covariance matrix of the stochastic process assuming the underlying variance is 1. The Choleski decomposition of the variance-covariance matrix will be used to generate the observations. In the second set of simulations the same procedure will be used but with the variance set equal to the estimate given in Table 2.9. i.e $\sigma^2 = 123.556$.

For sample sizes of 10, 30, 50 and 70 five Latin Hypercube Designs were generated for each set. For each Latin Hypercube the responses were generated using parameters of the fitted surrogate computer model. The methodology of Sacks *et al.* (1989b) was used to fit a surrogate computer model to the data and 1000 random check points over the experimental region were used so that the empirical mean square error could be calculated. The results are given in Table 5.1 and Figure 5.1.

The results show good agreement between the simulated values in Table 5.1 and the actual values as given in Table 3.5. In addition Table 5.1 indicates that simulating with $\sigma^2 = 1$ is sufficient since the average difference in log(EMSE)

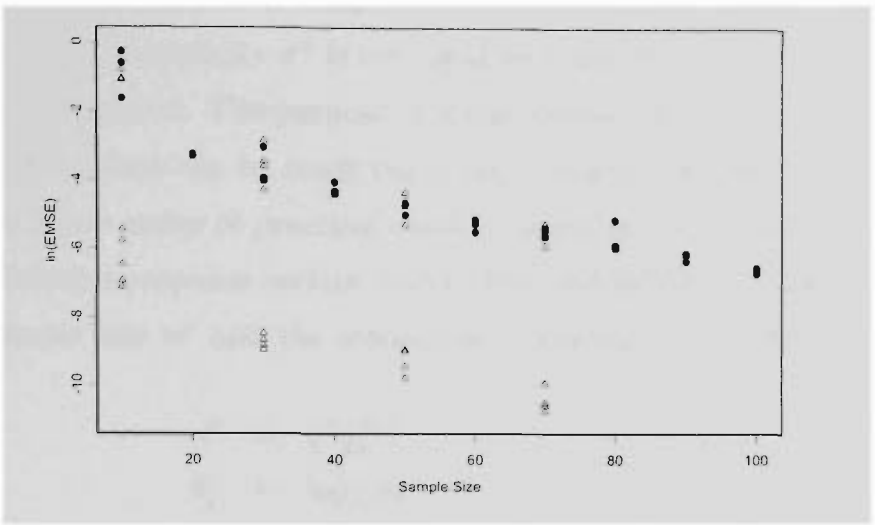

Figure 5.1: Comparison of Simulation of ASET-B with actual results. The actual results from Chapter 3 are denoted by filled in circles, while the simulated results are denoted by triangles. The top set refers to $\sigma^2 = 123.556$ and the bottom set refers to $\sigma^2 = 1$.

between the simulations with $\sigma^2 = 1$ and $\sigma^2 = 123.556$ are 5.48, 5.07, 4.65 and 4.94 for $n = 10, 30, 50$ and 70 respectively, which are quite consistent with $\ln(123.556) = 4.82$.

5.3 Design of the Experiment

The literature and previous chapters of the thesis have shown how a computer model can be approximated by a prediction equation. The quality of the prediction equation is measured in terms of the Empirical Mean Square Error (EMSE). This quantity is a function of the sample size n and the complexity of the computer model. The prediction equation is given in terms of the parameters

$$\begin{aligned} \theta &= (\theta_1, \theta_2, \ldots, \theta_d), \\ \mathbf{p} &= (p_1, p_2, \ldots, p_d) \end{aligned}$$

and σ^2.

In this chapter a simulation study is described where the parameters θ, $\mathbf{p}$, and n are regarded as inputs to a computer model and the $\ln(EMSE)$ is regarded as the output. For simplicity σ^2 is set equal to 1 and only the four dimensional case $d = 4$ is considered. The purpose of the simulation study is to examine what general considerations can be made regarding computer models as distinct from examining a whole series of practical cases as in the previous chapters.

A Box-Behnken response surface design (Box and Behnken, 1960) in the standardized sample size n^s and the standardized parameters θ^s and p^s was used, where

$$\begin{aligned} n^s &= \left(\tfrac{n-50}{30}\right) \\ \theta_i^s &= \log_{10}\theta_i + 2 \quad i = 1, \ldots, 4 \\ p_i^s &= \left(\tfrac{p_i - 1.75}{0.25}\right) \quad i = 1, \ldots, 4. \end{aligned}$$

In the Box-Behnken design, only the levels $-1, 0$ and 1 are used and therefore the levels for the input variables are given below:

Parameter	Level 1	Level 2	Level 3
n	20	50	80
θ	0.001	0.01	0.1
p	1.5	1.75	2

The design is given in Tables 5.3 to 5.5. A Box-Behnken design was used with the objective of approximating ln(EMSE) to a quadratic function of $n^s, \theta_1^s, \ldots, \theta_4^s$, $p_1^s, \ldots, p_4^s$. A Box-Behnken design was used since although an approximating function was required that applied over a range of values of n^s, θ_i^s, p_i^s, it was desired to make predictions for the ASET-B model where the p_i^s values are very close to 1. Using a Box-Behnken design would involve less extrapolation than using other response surface designs such as central composite designs.

n_s	θ_1^s	θ_2^s	θ_3^s	θ_4^s	p_1^s	p_2^s	p_3^s	p_4^s	$\ln(EMSE)$
−1	0	0	−1	0	0	−1	0	0	−6.689
−1	0	0	−1	0	0	1	0	0	−7.336
−1	0	0	1	0	0	−1	0	0	−5.701
−1	0	0	1	0	0	1	0	0	−6.175
1	0	0	−1	0	0	−1	0	0	−8.688
1	0	0	−1	0	0	1	0	0	−10.092
1	0	0	1	0	0	−1	0	0	−7.089
1	0	0	1	0	0	1	0	0	−7.831
0	−1	0	0	−1	0	0	−1	0	−8.409
0	−1	0	0	−1	0	0	1	0	−10.202
0	−1	0	0	1	0	0	−1	0	−7.216
0	−1	0	0	1	0	0	1	0	−7.296
0	1	0	0	−1	0	0	−1	0	−7.233
0	1	0	0	−1	0	0	1	0	−7.974
0	1	0	0	1	0	0	−1	0	−5.996
0	1	0	0	1	0	0	1	0	−6.235
0	0	−1	0	0	−1	0	0	−1	−7.975
0	0	−1	0	0	−1	0	0	1	−8.343
0	0	−1	0	0	1	0	0	−1	−8.738
0	0	−1	0	0	1	0	0	1	−10.021
0	0	1	0	0	−1	0	0	−1	−6.672
0	0	1	0	0	−1	0	0	1	−6.887
0	0	1	0	0	1	0	0	−1	−6.975
0	0	1	0	0	1	0	0	1	−7.401
−1	−1	−1	0	0	0	0	0	0	−7.132
−1	−1	1	0	0	0	0	0	0	−5.802
−1	1	−1	0	0	0	0	0	0	−6.139
−1	1	1	0	0	0	0	0	0	−4.833
1	−1	−1	0	0	0	0	0	0	−10.199
1	−1	1	0	0	0	0	0	0	−8.174
1	1	−1	0	0	0	0	0	0	−8.091
1	1	1	0	0	0	0	0	0	−6.542

Table 5.2: Runs 1–32 of the Box-Behnken experiment on the ASET-B Fire Model.

n_s	θ_1^s	θ_2^s	θ_3^s	θ_4^s	p_1^s	p_2^s	p_3^s	p_4^s	$\ln(EMSE)$
0	0	0	−1	−1	−1	0	0	0	−8.414
0	0	0	−1	−1	1	0	0	0	−10.250
0	0	0	−1	1	−1	0	0	0	−7.064
0	0	0	−1	1	1	0	0	0	−7.732
0	0	0	1	−1	−1	0	0	0	−7.321
0	0	0	1	−1	1	0	0	0	−7.815
0	0	0	1	1	−1	0	0	0	−6.112
0	0	0	1	1	1	0	0	0	−6.441
0	0	0	0	0	0	−1	−1	−1	−6.908
0	0	0	0	0	0	−1	−1	1	−7.972
0	0	0	0	0	0	−1	1	−1	−8.104
0	0	0	0	0	0	−1	1	1	−8.682
0	0	0	0	0	0	1	−1	−1	−7.804
0	0	0	0	0	0	1	−1	1	−8.297
0	0	0	0	0	0	1	1	−1	−8.884
0	0	0	0	0	0	1	1	1	−10.359
−1	0	0	0	−1	0	0	0	−1	−6.689
−1	0	0	0	−1	0	0	0	1	−7.247
−1	0	0	0	1	0	0	0	−1	−5.073
−1	0	0	0	1	0	0	0	1	−6.102
1	0	0	0	−1	0	0	0	−1	−9.155
1	0	0	0	−1	0	0	0	1	−9.551
1	0	0	0	1	0	0	0	−1	−7.111
1	0	0	0	1	0	0	0	1	−9.038
0	0	−1	−1	0	0	0	−1	0	−8.946
0	0	−1	−1	0	0	0	1	0	−9.532
0	0	−1	1	0	0	0	−1	0	−6.978
0	0	−1	1	0	0	0	1	0	−8.587
0	0	1	−1	0	0	0	−1	0	−6.503
0	0	1	−1	0	0	0	1	0	−7.208
0	0	1	1	0	0	0	−1	0	−5.807
0	0	1	1	0	0	0	1	0	−6.765

Table 5.3: Runs 33 64 of the Box-Behnken experiment on the ASET-B Fire Model.

n_s	θ_1^s	θ_2^s	θ_3^s	θ_4^s	p_1^s	p_2^s	p_3^s	p_4^s	$\ln(EMSE)$
0	−1	0	0	0	−1	−1	0	0	−8.307
0	−1	0	0	0	−1	1	0	0	−8.721
0	−1	0	0	0	1	−1	0	0	−8.749
0	−1	0	0	0	1	1	0	0	−9.943
0	1	0	0	0	−1	−1	0	0	−6.599
0	1	0	0	0	−1	1	0	0	−6.351
0	1	0	0	0	1	−1	0	0	−7.400
0	1	0	0	0	1	1	0	0	−8.768
−1	0	0	0	0	−1	0	−1	0	−6.341
−1	0	0	0	0	−1	0	1	0	−7.070
−1	0	0	0	0	1	0	−1	0	−6.664
−1	0	0	0	0	1	0	1	0	−7.617
1	0	0	0	0	−1	0	−1	0	−7.958
1	0	0	0	0	−1	0	1	0	−8.883
1	0	0	0	0	1	0	−1	0	−8.869
1	0	0	0	0	1	0	1	0	−10.189
0	−1	0	−1	0	0	0	0	−1	−8.650
0	−1	0	−1	0	0	0	0	1	−10.065
0	−1	0	1	0	0	0	0	−1	−7.106
0	−1	0	1	0	0	0	0	1	−7.747
0	1	0	−1	0	0	0	0	−1	−6.538
0	1	0	−1	0	0	0	0	1	−8.242
0	1	0	1	0	0	0	0	−1	−6.106
0	1	0	1	0	0	0	0	1	−6.068
0	0	−1	0	−1	0	−1	0	0	−8.456
0	0	−1	0	−1	0	1	0	0	−9.171
0	0	−1	0	1	0	−1	0	0	−7.232
0	0	−1	0	1	0	1	0	0	−7.825
0	0	1	0	−1	0	−1	0	0	−6.927
0	0	1	0	−1	0	1	0	0	−9.033
0	0	1	0	1	0	−1	0	0	−5.947
0	0	1	0	1	0	1	0	0	−6.195

Table 5.4: Runs 65 96 of the Box-Behnken experiment on the ASET-B Fire Model.

n_s	θ_1^s	θ_2^s	θ_3^s	θ_4^s	p_1^s	p_2^s	p_3^s	p_4^s	$\ln(EMSE)$
−1	0	0	−1	0	0	−1	0	0	−6.023
−1	0	0	−1	0	0	1	0	0	−7.365
−1	0	0	1	0	0	−1	0	0	−5.555
−1	0	0	1	0	0	1	0	0	−5.692
1	0	0	−1	0	0	−1	0	0	−8.573
1	0	0	−1	0	0	1	0	0	−9.941
1	0	0	1	0	0	−1	0	0	−7.549
1	0	0	1	0	0	1	0	0	−8.169
0	−1	0	0	−1	0	0	−1	0	−8.377
0	−1	0	0	−1	0	0	1	0	−10.231
0	−1	0	0	1	0	0	−1	0	−7.516
0	−1	0	0	1	0	0	1	0	−7.729
0	1	0	0	−1	0	0	−1	0	−7.323
0	1	0	0	−1	0	0	1	0	−7.668
0	1	0	0	1	0	0	−1	0	−5.759
0	1	0	0	1	0	0	1	0	−6.132
0	0	−1	0	0	−1	0	0	−1	−7.470
0	0	−1	0	0	−1	0	0	1	−8.388
0	0	−1	0	0	1	0	0	−1	−8.826
0	0	−1	0	0	1	0	0	1	−10.142
0	0	1	0	0	−1	0	0	−1	−6.269
0	0	1	0	0	−1	0	0	1	−7.125
0	0	1	0	0	1	0	0	−1	−6.843
0	0	1	0	0	1	0	0	1	−7.382
0	0	0	0	0	0	0	0	0	−7.991
0	0	0	0	0	0	0	0	0	−6.902
0	0	0	0	0	0	0	0	0	−8.601
0	0	0	0	0	0	0	0	0	−7.742
0	0	0	0	0	0	0	0	0	−8.483
0	0	0	0	0	0	0	0	0	−8.213
0	0	0	0	0	0	0	0	0	−8.045
0	0	0	0	0	0	0	0	0	−8.147
0	0	0	0	0	0	0	0	0	−7.963
0	0	0	0	0	0	0	0	0	−7.578

Table 5.5: Runs 97 130 of the Box-Behnken experiment on the ASET-B Fire Model.

The second order model is

$$
\begin{aligned}
&\ln(\text{EMISE}) \\
&= b_0 + b_1 n^s + b_2\theta_1^s + b_3\theta_2^s + b_4\theta_3^s + b_5\theta_4^s + b_6 p_1^s + b_7 p_2^s + b_8 p_3^s + b_9 p_4^s \\
&\quad + b_{11}(n^s)^2 + b_{22}(\theta_1^s)^2 + b_{33}(\theta_2^s)^2 + b_{44}(\theta_3^s)^2 + b_{55}(\theta_4^s)^2 \\
&\quad + b_{66}(p_1^s)^2 + b_{77}(p_2^s)^2 + b_{88}(p_3^s)^2 + b_{99}(p_4^s)^2 \\
&\quad + b_{12}(n^s\theta_1^s) + b_{13}(n^s\theta_2^s) + b_{14}(n^s\theta_3^s) + b_{15}(n^s\theta_4^s) \\
&\quad + b_{16}(n^s p_1^s) + b_{17}(n^s p_2^s) + b_{18}(n^s p_3^s) + b_{19}(n^s p_4^s) \\
&\quad + b_{23}(\theta_1^s\theta_2^s) + b_{24}(\theta_1^s\theta_3^s) + b_{25}(\theta_1^s\theta_4^s) \\
&\quad + b_{26}(\theta_1^s p_1^s) + b_{27}(\theta_1^s p_2^s) + b_{28}(\theta_1^s p_3^s) + b_{29}(\theta_1^s p_4^s) \\
&\quad + b_{34}(\theta_2^s\theta_3^s) + b_{35}(\theta_2^s\theta_4^s) + b_{36}(\theta_2^s p_1^s) + b_{37}(\theta_2^s p_2^s) + b_{38}(\theta_2 p_3^s) + b_{39}(\theta_2 p_4^s) \\
&\quad + b_{45}(\theta_3^s\theta_4^s) + b_{46}(\theta_3^s p_1^s) + b_{47}(\theta_3^s p_2^s) + b_{48}(\theta_3^s p_3^s) + b_{49}(\theta_3^s p_4^s) \\
&\quad + b_{56}(\theta_4^s p_1^s) + b_{57}(\theta_4^s p_2^s) + b_{58}(\theta_4^s p_3^s) + b_{59}(\theta_4^s p_4^s) \\
&\quad + b_{67}(p_1^s p_2^s) + b_{68}(p_1^s p_3^s) + b_{69}(p_1^s p_4^s) + b_{78}(p_2^s p_3^s) + b_{79}(p_2^s p_4^s) + b_{89}(p_3^s p_4^s) \\
&= b_0 + b_1 n^s + b_2(\theta_1^s + \theta_2^s + \theta_3^s + \theta_4^s) + b_6(p_1^s + p_2^s + p_3^s + p_4^s) \\
&\quad + b_{11}(n^s)^2 + b_{22}\left[(\theta_1^s)^2 + (\theta_2^s)^2 + (\theta_3^s)^2 + (\theta_4^s)^2\right] \\
&\quad + b_{66}\left[(p_1^s)^2 + (p_2^s)^2 + (p_3^s)^2 + (p_4^s)^2\right] \\
&\quad + b_{12}\left[n^s(\theta_1^s + \theta_2^s + \theta_3^s + \theta_4^s)\right] \\
&\quad + b_{16}\left[n^s(p_1^s + p_2^s + p_3^s + p_4^s)\right] \\
&\quad + b_{23}\left[(\theta_1^s\theta_2^s) + (\theta_1^s\theta_3^s) + (\theta_1^s\theta_4^s) + (\theta_2^s\theta_3^s) + (\theta_2^s\theta_4^s) + (\theta_3^s\theta_4^s)\right] \\
&\quad + b_{26}\left[(\theta_1^s p_1^s) + (\theta_2^s p_2^s) + (\theta_3^s p_3^s) + (\theta_4^s p_4^s)\right] \\
&\quad + b_{27}\left[\theta_1^s(p_2^s + p_3^s + p_4^s) + \theta_2^s(p_1^s + p_3^s + p_4^s) + \theta_3^s(p_1^s + p_2^s + p_4^s) + \theta_4^s(p_1^s + p_2^s + p_3^s)\right] \\
&\quad + b_{67}\left[p_1^s p_2^s + p_1^s p_3^s + p_1^s p_4^s + p_2^s p_3^s + p_2^s p_4^s + p_3^s p_4^s\right]
\end{aligned}
$$

since, by symmetry, many of the parameters will be the same. As there is no design distinction between input parameters, the effect of all θ and p should be the same.

To interpret the results, the average, main effects, marginal effects, two-factor interaction effects and joint effects were determined. For a quadratic model involving d factors, say $x_1, x_2, \ldots, x_d$ all over the range $-1 < x_i < 1$ the average

```
> summary(fourdcasefit, cor = F)

Coefficients:
```

	Value	Std.Error	T-Stat
(Intercept)	-8.0665	0.0900	-89.6048
ns	-1.1062	0.0450	-24.5754
I(th1s+th2s+th3s+th4s)	0.7725	0.0225	34.3237
I(p1s+p2s+p3s+p4s)	-0.4421	0.0225	-19.6438
I(ns^2)	0.3886	0.0581	6.6868
I(th1s^2+th2s^2+th3s^2+th4s^2)	0.2950	0.0375	7.8765
I(p1s^2+p2s^2+p3s^2+p4s^2)	-0.0955	0.0375	-2.5494
I(ns*(th1s+th2s+th3s+th4s))	0.1169	0.0450	2.5982
I(ns*p1s+ ... +ns*p4s)	-0.1095	0.0450	-2.4327
I(th1s*p1s+ ... +th4s*p4s)	-0.1899	0.0503	-3.7743
I(th1s*(p2s+p3s+p4s)+th4s*(p1s+p2s+p3s))	0.1731	0.0244	7.0906
I(th1s*th2s+th1s*th3s+th3s*th4s)	-0.0884	0.0380	-2.3226
I(p1s*p2s+p1s*p3s+p3s*p4s)	-0.1121	0.0380	-2.9479

```
Residual standard error: 0.2847 on 117 degrees of freedom
Multiple R-Squared: 0.9536 F-statistic: 200.2 on 12 and 117
degrees of freedom, the p-value is 0
```

Table 5.6: Results for the Box-Behnken design on the ASET-B Computer Model

over the region is given by

$$\begin{aligned}\mu_0 &= \frac{1}{2^d}\int_{x_i\in[-1,1]}\cdots\int\left(a_0+\sum_i a_i x_i+\sum_i a_{ii}x_i^2+\sum_{i<j}a_{ij}x_i x_j\right)\prod dx_i \\ &= a_0+\frac{1}{3}\left(\sum_i a_{ii}\right).\end{aligned}$$

Similarly, the main effect of x_k over the region is given by

$$\begin{aligned}\mu_k(x_k) &= \frac{1}{2^{d-1}}\int_{x_{i\neq k}\in[-1,1]}\cdots\int\left(a_0+\sum_i a_i x_i+\sum_i a_{ii}x_i^2+\sum_{i<j}a_{ij}x_i x_j\right)\prod_{i\neq k}dx_i-\mu_0 \\ &= a_0+a_k x_k+a_{kk}x_k^2+\frac{1}{3}\left(\sum_{i\neq k}a_{ii}\right)-\mu_0 \\ &= a_k x_k+a_{kk}x_k^2-\frac{1}{3}a_{kk}.\end{aligned}$$

The marginal effect of x_k is given by

$$\begin{aligned}&\mu_0+\mu_k(x_k) \\ &\quad= a_0+a_k x_k+a_{kk}x_k^2+\frac{1}{3}\left(\sum_{i\neq k}a_{ii}\right).\end{aligned}$$

Also, the interaction effect of x_k and x_l is given by

$$\begin{aligned}\mu_{kl}(x_k,x_l) &= \frac{1}{2^{d-2}}\int_{x_{i\neq k,l}\in[-1,1]}\cdots\int\left(a_0+\sum_i a_i x_i+\sum_i a_{ii}x_i^2+\sum_{i<j}a_{ij}x_i x_j\right)\prod_{i\neq k,l}dx_i \\ &\quad-\mu_k(x_k)-\mu_l(x_l)-\mu_0 \\ &= a_0+a_k x_k+a_l x_l+a_{kk}x_k^2+a_{ll}x_l^2+a_{kl}x_k x_l \\ &\quad+\frac{1}{3}\left(\sum_{i\neq j,k}a_{ii}\right)-\mu_k(x_k)-\mu_l(x_l)-\mu_0 \\ &= a_{kl}x_k x_l.\end{aligned}$$

Finally, the joint effect of x_k and x_l is given by

$$\begin{aligned}&\mu_0+\mu_k(x_k)+\mu_l(x_l)+\mu_{kl}(x_k,x_l) \\ &\quad= a_0+\frac{1}{3}\left(\sum_{i\neq j,k}a_{ii}\right)+a_k x_k+a_l x_l+a_{kk}x_k^2+a_{ll}x_l^2+a_{kl}x_k x_l.\end{aligned}$$

Applying the above expressions to the results given in Table 5.6 the estimated average, main effects, marginal effects, two factor interaction effects and joint effects were calculated.

Figure 5.2 gives the main effects of θ, p and n. The graph shows that as θ increases the log of the empirical mean square error increases, while as p and n increase the the log of the empirical mean square error decreases.

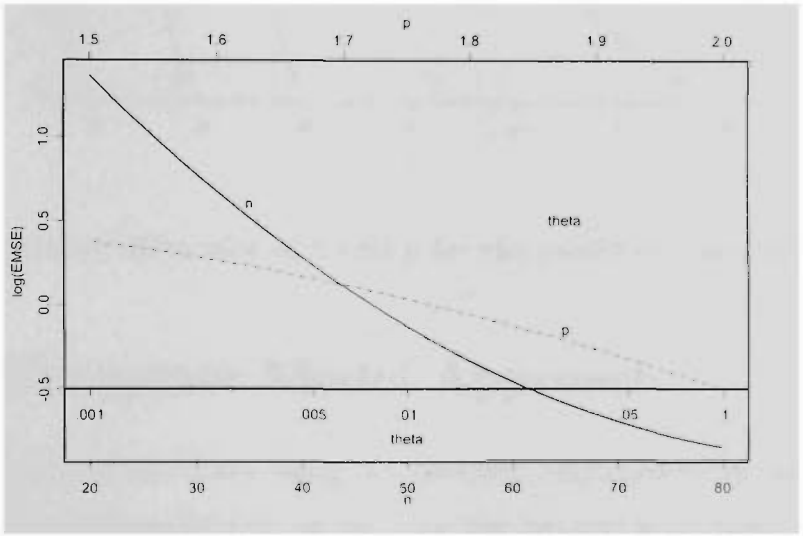

Figure 5.2: Main effects plots for the model of computer models.

Figure 5.4 gives the joint effect of n and θ in terms of ln(EMSE). The graph shows that the effects of n and θ are approximately independent. Figure 5.3 gives the joint effect of p and n. The effect of n on ln(EMSE) is greater at high values of p than at low values of p.

Figure 5.5 gives the joint effect of θ_1 and p_1 in terms of log(EMSE). The effect of θ_1 on ln(EMSE) is greater when p_1 is low than when p_2 is high. Figure 5.6 gives the joint effect of θ_1 and p_2. This graph shows a very similar pattern to that for θ_1 and p_1.

Figure 5.7 gives the joint effect of θ_1 and θ_2. The graph shows that the effects are approximately independent. Finally, Figure 5.8 gives the joint effect of p_1 and p_2. The graph shows that the effects are also approximately independent.

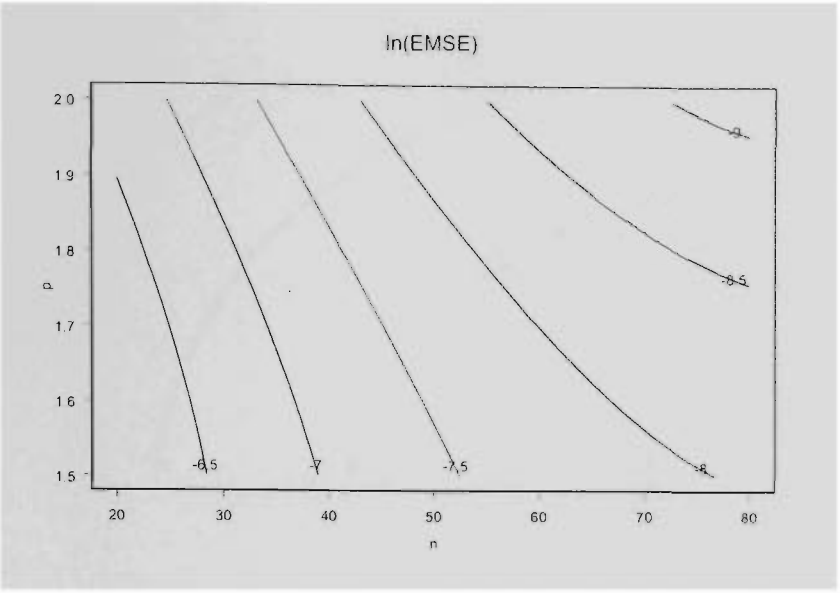

Figure 5.3: Joint effect plot of n and p for the model of computer models.

5.4 A Computer Model Approach

Rather than analyse the data using a quadratic approximation, an alternative approach is to model the output, in this case the natural logarithm of the EMSE, as a function of the input parameters: n, θ and p. An important difference is that we want to *smooth* the response and not interpolate since different simulations give different responses.

In this case the output is modelled as a realisation of a stochastic process with mean β and variance-covariance matrix

$$\sigma_1^2 R_D$$

but with an added independent measurement error σ_0^2. The model can be fitted using maximum likelihood as in previous chapters. However if, for example, there were four replications for each run of the design, this is made much easier

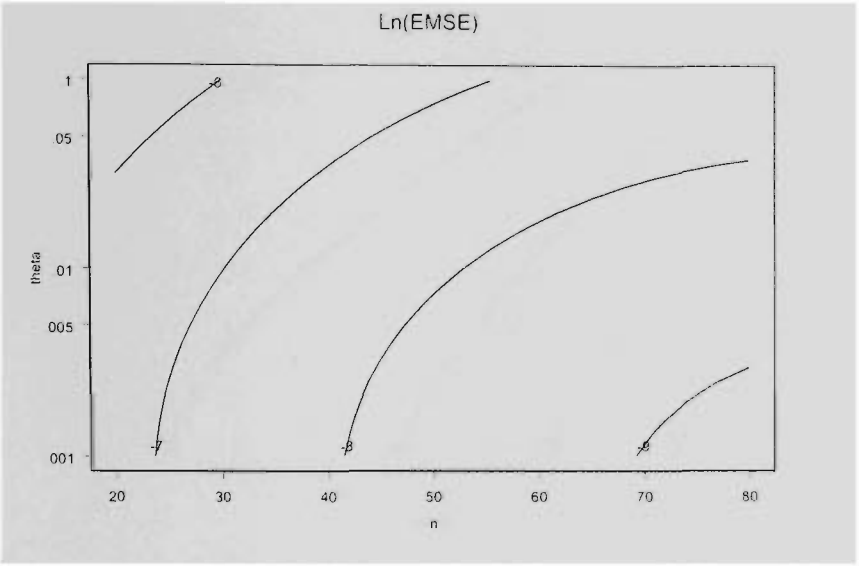

Figure 5.4: Joint Effects of n and θ for the model of computer models.

by writing the complete R_D matrix as:

$$\begin{pmatrix} A & B & B & B \\ B & A & B & B \\ B & B & A & B \\ B & B & B & A \end{pmatrix}$$

where

$$\begin{aligned} A &= (1-\gamma)I + \gamma R_{\mathcal{E}} \\ B &= \gamma R_{\mathcal{E}} \\ \gamma &= \frac{\sigma_1^2}{\sigma_0^2 + \sigma_1^2}. \end{aligned}$$

In the above equation R_D is the variance-covariance matrix of the entire design while $R_{\mathcal{E}}$ is the variance-covariance matrix of one of the replicates. The parameter γ is a smoothing parameter between 0 and 1 which is estimated along with the parameters that make up $R_{\mathcal{E}}$. The closer γ is to 1 the closer the prediction is to an interpolation.

When calculating the likelihood the inverse of the R_D matrix can be calculated

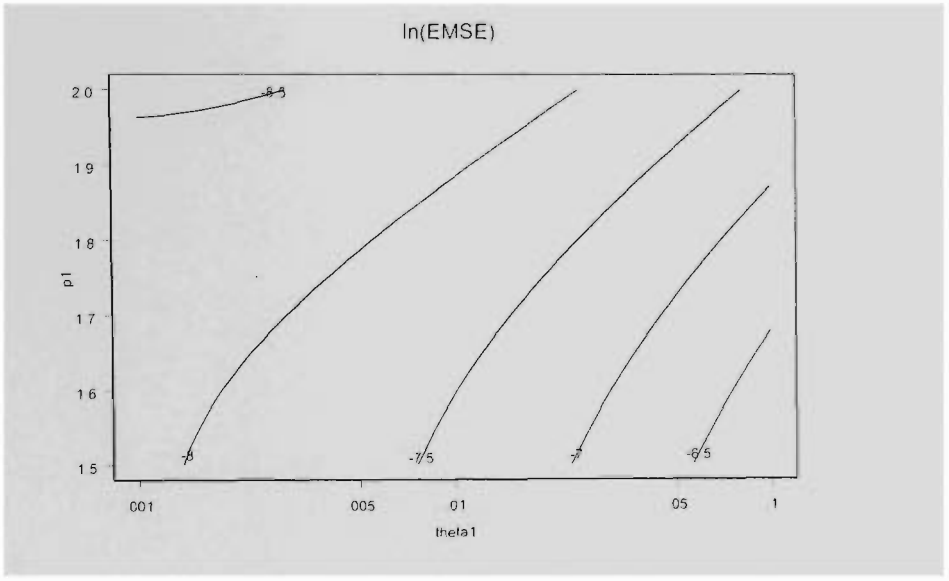

Figure 5.5: Joint effect plot of θ and p for a one-dimensional computer model.

as

$$\begin{pmatrix} C & D & D & D \\ D & C & D & D \\ D & D & C & D \\ D & D & D & C \end{pmatrix}$$

where

$$\begin{aligned} C &= A^{-1} + A^{-1}B(A + 2B - BA^{-1}B)^{-1}BA^{-1}/3 \\ D &= -(A + 2B - BA^{-1}B)^{-1}BA^{-1}/3. \end{aligned}$$

Similarly, the determinant of $R_{\mathcal{E}}$ is required. If say a 50 run LHD had been used with 4 replicates then the determinant is conveniently calculated as

$$\begin{aligned} |R_D| &= |A - B|^3|A + 3B| \\ &= (1 - \gamma)^{150}|A + 3B|. \end{aligned}$$

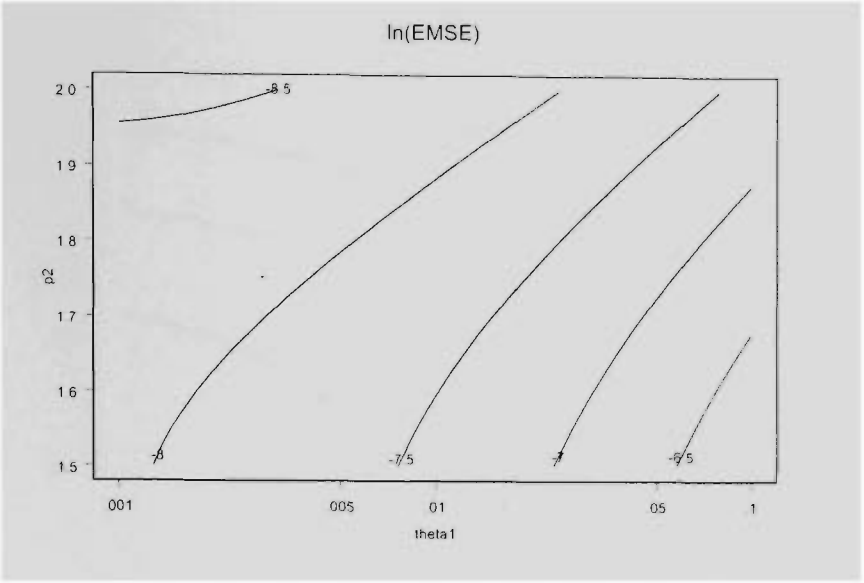

Figure 5.6: Joint effect plot of θ_1 and p_2 for the model of computer models.

5.5 Some Applications of the Results

The model developed in the previous section used $\sigma^2 = 1$. For other values of σ the marginal effect of n^s, for example, is given by

$$\ln(\text{EMSE}) = \log(\sigma^2) - \frac{1}{3}(b_{22} + b_{66}) - b_1 n^s - b_{11}(n^s)^2.$$

Figure 5.9 gives the results of Table 3.5, the predictions for the fitted quadratic model given in Table 5.6 with the parameters as in Table 2.9, and the marginal effect with $\sigma^2 = 123.556$.

The agreement between the results of Table 3.5 and the predictions from the fitted quadratic model is quite good. On the other hand the marginal effect is biased but also parallel to the pattern exhibited by the data.

5.6 Conclusions

In this chapter an experiment on computer models has been conducted. This experiment was concerned with establishing a relationship between the Empirical Mean Square Error and the parameters of a four dimensional computer model.

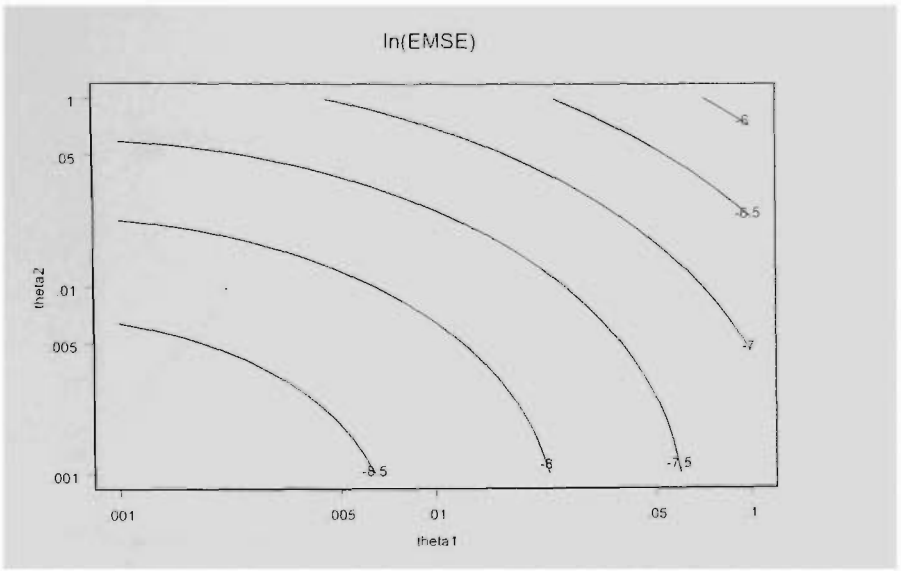

Figure 5.7: Joint effect plot of θ_1 and θ_2 for the model of computer models.

$\theta_1, \ldots, \theta_4, p_1, \ldots, p_4$ and the sample size n.

A Box-Behnken design was used and a second order response surface model was fitted to the data. The form of the model was simplified to take into account symmetries in the parameters. The quadratic model fitted to the ASET-B computer code studied in Chapter 3 is a good fit (residual standard error=0.2847). and the model accounted for over 95% of the variation in ln(EMSE) with R-Squared = 0.9536 (see Table 5.6).

Expressions for the mean, main effects, marginal effects, two factor interaction effects and joint effects were obtained and applied to the ASET-B model.

If knowledge of the likely values of θ, p and σ^2 are available, the response surface model is shown, at least for ASET-B, to give an approximate relationship between ln(EMSE) and the sample size. If such knowledge is unavailable then the marginal effect of n on ln(EMSE) can be used to approximate the relationship up to an additive constant.

Although the results are useful better results might be obtained by using the machinery of computer models to develop a model but smoothing rather than interpolating the responses.

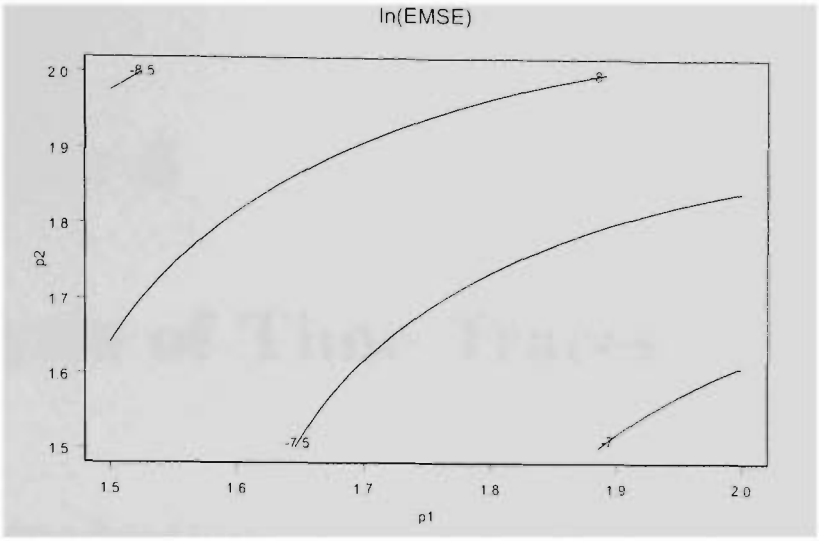

Figure 5.8: Joint effect plot of p_1 and p_2 for the model of computer models.

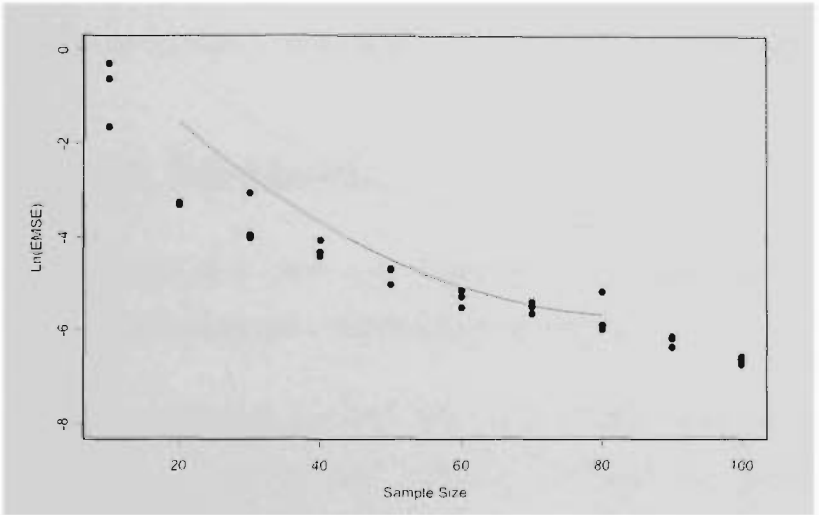

Figure 5.9: Comparison of results for the ASET-B model (data points) and the predicted response from the response surface model (solid line) and the predictions from the marginal model (dotted line).

Chapter 6

Analysis of Time Traces

6.1 Introduction

Many computer models provide a trace of responses at various values of a time parameter. For example, the ASET-B program provides the height of the smoke layer at intervals of 5 seconds.

Although such time trace responses are very commonly used, the literature has focused on computer experiments with a univariate response. In this chapter consideration will be on how a response as a function of time could be analysed.

6.2 Possible Methods

When the data consists of a trace over time for the various input conditions a number of ways of analysing the data suggest themselves:

1. *Do a separate calculation for each time point.* This would give estimates of $\theta_1, \ldots, \theta_d$ and $p_1, \ldots, p_d$ for each value of time that has been investigated. Hence estimates of the response at each of the times can be generated using the appropriate values of the θs and ps in the linear predictor. To get estimates of the response at intermediate times a cubic smoothing spline of the predictions can be made.

2. *Regard time as an additional input parameter.* This would give rise to two additional parameters θ_{d+1} and p_{d+1} describing the correlation as a function of time. The data vector from an $n-$run Latin Hypercube design with T times would be of length nT.

Both methods will be developed and applied in turn to the ASET-B model. The advantages and disadvantages of each method will be discussed.

Assuming that a time average gives a complete representation of the full ensemble in the second method described above, this system can be considered as a special case of ergodic properties of particular interest - Ergodicity.

Ergodicity

Ergodicity is an attribute of a stochastic process. The foundation of the Markov chain theory is the Ergodicity Theorem. It establishes the conditions under which a Markov chain can be analysed to determine its steady state behaviour. In general it is a process that leads to the probability of a limiting form that is independent of the initial conditions. In the proposed second method above, say time as an additional input parameter, it is about the time average property of the process that carries an implicit ergodicity [see Gray (1987), for more details].

It is common to assume ergodicity, where there is interchangeability of time averages with the ensemble average. In practical terms, this means measuring the impulse responses at different times and then computing the time delay correlation function. In this scenario, evaluation of ergodicity as a feature of dynamic behaviour is important, not opposing the assumption facilitating the use of a probability model (Sposito, 1997). A paper by Pribadi *et al.* (2001) on a reduction method for Markov chains proved the importance of preserving the property of ergodicity. This approach is especially useful in the area of simulation and performance estimation.

6.2.1 Separate Calculation for a Number of Time Points

A 50 run LHD was generated. For each run the ASET-B program was used to generate the height of the smoke layer at 5 second intervals up to 100 seconds. Figure 6.1 gives the generated traces. Table 6.1 and Table 6.2 give the input variables for the 50 run LHD. Table 6.3. Table 6.4. Table 6.5. Table 6.6. Table 6.7 and Table 6.8 give the heights of the smoke layer at 5 second intervals up to 100 seconds for runs 1 to 50 of the same LHD.

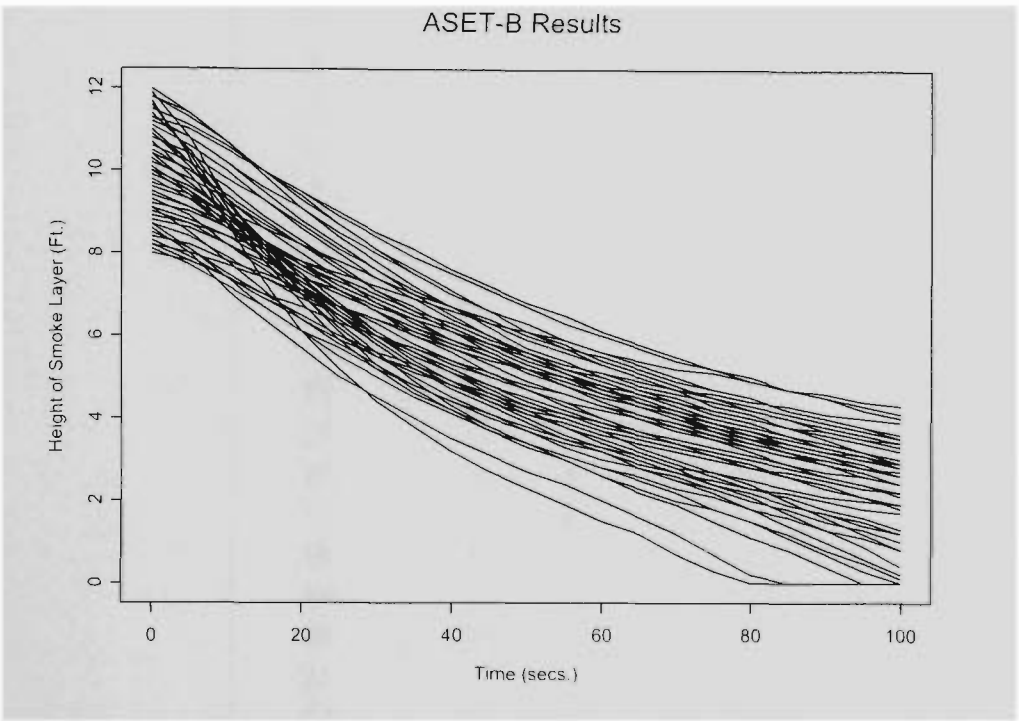

Figure 6.1: Generated Time Traces for the ASET-B Latin Hypercube Experiment

For each time period a separate computer model was fitted using the methodology of Sacks *et al.* (1989b), giving values of θ and p for each time period. Figure 6.2 and Figure 6.3 give plots of the values of θ and p. The graphs show that there is considerable variation in the θ and p values and that these values are not smooth.

For a Heat Loss Fraction of 0.85, Fire Height of 2.5 ft, Room Ceiling Height of 9 ft and Room Floor Area of 150 sq. ft, predictions of the smoke layer height at

Run No.	$49x_1$	$49x_2$	$49x_3$	$49x_4$
1	−45	−25	15	27
2	−5	−45	−1	−7
3	43	−47	−47	−5
4	−1	−19	47	−43
5	37	−29	−23	−25
6	−13	43	29	17
7	−47	−35	35	−29
8	−7	−41	9	−15
9	−3	−27	23	−21
10	−37	9	5	19
11	19	23	−49	−1
12	3	47	−21	−9
13	−41	−31	31	37
14	49	45	−29	3
15	−15	25	−11	47
16	−39	−39	−25	33
17	−9	29	−13	−3
18	17	−13	27	21
19	31	1	19	29
20	25	17	3	−31
21	7	3	49	11
22	47	39	−27	41
23	−19	7	−33	−41
24	−27	37	33	49
25	−31	5	11	13

Table 6.1: Scaled LHD points for ASET-B Input Variables. [Runs 1 to 25 of $N = 50$].

Run No.	$49x_1$	$49x_2$	$49x_3$	$49x_4$
26	−33	−43	41	−49
27	39	−23	−3	9
28	45	−17	−5	−11
29	−17	−15	−7	23
30	33	15	37	45
31	−23	−33	1	15
32	−43	33	−17	−27
33	13	35	−41	−35
34	21	19	−31	−13
35	41	−37	−9	35
36	15	13	−45	−17
37	−29	11	21	−37
38	−11	−49	7	1
39	5	−1	45	31
40	1	41	−15	5
41	−35	−5	43	39
42	23	−11	39	−47
43	35	−7	−37	−19
44	27	−3	−19	43
45	−21	21	−43	25
46	−25	27	25	−39
47	29	31	−39	7
48	−49	−9	−35	−45
49	11	−21	17	−33
50	9	49	13	−23

Table 6.2: Scaled LHD points for ASET-B Input Variables. [Runs 26 to 50 of $N = 50$].

Run No.	0s	5s	10s	15s	20s	25s	30s
1	10.60	10.20	9.60	8.90	8.30	7.70	7.10
2	10.00	9.40	8.60	7.90	7.10	6.50	5.90
3	8.10	7.70	7.20	6.70	6.10	5.70	5.20
4	11.90	10.80	9.30	8.10	7.00	6.10	5.30
5	9.10	8.50	7.80	7.10	6.40	5.80	5.30
6	11.20	10.80	10.20	9.60	9.10	8.50	8.00
7	11.40	10.50	9.30	8.20	7.20	6.30	5.60
8	10.40	9.70	8.90	8.00	7.20	6.50	5.80
9	10.90	10.20	9.20	8.30	7.40	6.70	6.00
10	10.20	9.80	9.30	8.70	8.20	7.70	7.10
11	8.00	7.80	7.40	7.00	6.70	6.30	6.00
12	9.10	8.80	8.40	7.90	7.40	7.00	6.60
13	11.30	10.80	10.20	9.50	8.80	8.20	7.60
14	8.80	8.60	8.20	7.80	7.40	7.10	6.70
15	9.60	9.30	9.00	8.60	8.20	7.80	7.40
16	9.00	8.70	8.20	7.70	7.20	6.70	6.30
17	9.50	9.10	8.60	8.10	7.60	7.10	6.60
18	11.10	10.60	10.00	9.30	8.60	8.00	7.40
19	10.80	10.40	9.80	9.30	8.70	8.10	7.60
20	10.10	9.50	8.70	7.90	7.20	6.60	6.00
21	12.00	11.40	10.70	9.90	9.10	8.40	7.80
22	8.90	8.70	8.40	8.10	7.80	7.50	7.20
23	8.70	8.10	7.40	6.70	6.10	5.50	5.00
24	11.30	11.00	10.50	10.00	9.50	9.00	8.50
25	10.40	10.00	9.40	8.80	8.20	7.60	7.10

Table 6.3: Scaled LHD points for ASET-B Input Variables. [Runs 1 to 25 of $N = 50$] and corresponding Egress Times from 0 to 30 seconds in five second intervals.

Run No.	0s	5s	10s	15s	20s	25s	30s
26	11.70	10.30	8.70	7.30	6.10	5.20	4.40
27	9.90	9.50	8.90	8.20	7.60	7.10	6.60
28	9.80	9.30	8.60	7.90	7.30	6.70	6.10
29	9.70	9.40	8.90	8.30	7.80	7.20	6.80
30	11.50	11.10	10.60	10.00	9.50	8.90	8.40
31	10.00	9.60	9.00	8.30	7.70	7.10	6.60
32	9.30	8.90	8.20	7.60	7.00	6.50	6.00
33	8.30	7.90	7.40	6.90	6.40	5.90	5.50
34	8.70	8.40	7.90	7.40	6.90	6.50	6.00
35	9.60	9.30	8.80	8.30	7.70	7.20	6.80
36	8.20	7.80	7.40	6.90	6.50	6.00	5.60
37	10.90	10.10	9.00	8.10	7.20	6.40	5.80
38	10.30	9.70	9.00	8.20	7.50	6.80	6.10
39	11.80	11.40	10.70	10.00	9.30	8.70	8.10
40	9.40	9.10	8.70	8.20	7.70	7.30	6.90
41	11.80	11.30	10.70	10.00	9.40	8.70	8.10
42	11.60	10.40	9.10	7.80	6.80	5.90	5.20
43	8.50	8.10	7.50	7.00	6.50	6.00	5.50
44	9.20	9.00	8.60	8.20	7.80	7.40	7.00
45	8.20	8.00	7.70	7.40	7.00	6.70	6.40
46	11.00	10.20	9.20	8.20	7.40	6.60	6.00
47	8.40	8.20	7.80	7.50	7.10	6.70	6.40
48	8.60	8.00	7.10	6.40	5.70	5.00	4.50
49	10.70	9.90	8.80	7.80	7.00	6.20	5.50
50	10.50	10.00	9.30	8.60	8.00	7.40	6.80

Table 6.4: Scaled LHD points for ASET-B Input Variables. [Runs 26 to 50 of $N = 50$] and corresponding Egress Times from 0 to 30 seconds in five second intervals.

Run No.	$35s$	$40s$	$45s$	$50s$	$55s$	$60s$	$65s$
1	6.60	6.10	5.60	5.20	4.80	4.40	4.10
2	5.30	4.80	4.40	4.00	3.60	3.30	3.00
3	4.80	4.40	4.10	3.80	3.50	3.30	3.10
4	4.70	4.20	3.70	3.30	3.00	2.60	2.30
5	4.80	4.40	4.00	3.70	3.40	3.20	2.90
6	7.50	7.00	6.60	6.30	5.90	5.60	5.30
7	4.90	4.30	3.80	3.40	3.00	2.60	2.20
8	5.20	4.70	4.30	3.90	3.50	3.20	2.80
9	5.40	4.90	4.40	4.00	3.60	3.30	3.00
10	6.70	6.20	5.80	5.40	5.10	4.80	4.40
11	5.70	5.40	5.10	4.80	4.60	4.30	4.10
12	6.20	5.90	5.50	5.20	5.00	4.70	4.50
13	7.00	6.50	6.00	5.50	5.10	4.70	4.40
14	6.40	6.10	5.80	5.60	5.30	5.10	4.90
15	7.00	6.70	6.40	6.00	5.70	5.50	5.20
16	5.80	5.40	5.00	4.70	4.30	4.00	3.70
17	6.20	5.80	5.50	5.10	4.80	4.50	4.30
18	6.90	6.40	6.00	5.60	5.20	4.90	4.60
19	7.10	6.70	6.30	5.90	5.60	5.30	5.00
20	5.60	5.10	4.80	4.40	4.10	3.90	3.60
21	7.20	6.70	6.20	5.70	5.40	5.00	4.70
22	6.90	6.60	6.30	6.10	5.90	5.60	5.40
23	4.50	4.10	3.70	3.40	3.10	2.80	2.50
24	8.10	7.60	7.20	6.80	6.50	6.10	5.80
25	6.60	6.10	5.70	5.30	5.00	4.60	4.30

Table 6.5: Scaled LHD points for ASET-B Input Variables. [Runs 1 to 25 of $N = 50$] and corresponding Egress Times from 35 to 65 seconds in five second intervals.

Run No.	$35s$	$40s$	$45s$	$50s$	$55s$	$60s$	$65s$
26	3.80	3.20	2.70	2.30	1.90	1.50	1.20
27	6.10	5.70	5.30	4.90	4.60	4.30	4.00
28	5.60	5.20	4.80	4.50	4.20	3.90	3.70
29	6.30	5.90	5.50	5.10	4.80	4.40	1.10
30	7.90	7.50	7.10	6.70	6.30	6.00	5.70
31	6.00	5.60	5.10	4.70	4.40	4.00	3.70
32	5.50	5.10	4.70	4.30	4.00	3.70	3.30
33	5.20	4.80	4.50	4.20	4.00	3.70	3.50
34	5.70	5.30	5.00	4.70	4.40	4.20	4.00
35	6.30	5.90	5.50	5.20	4.90	4.60	4.30
36	5.30	4.90	4.60	4.30	4.10	3.80	3.60
37	5.20	4.70	4.30	3.90	3.50	3.20	2.80
38	5.60	5.10	4.60	4.20	3.80	3.50	3.20
39	7.50	7.00	6.60	6.10	5.80	5.40	5.10
40	6.50	6.20	5.80	5.50	5.20	5.00	4.70
41	7.60	7.10	6.60	6.10	5.70	5.30	5.00
42	4.60	4.20	3.80	3.40	3.10	2.80	2.60
43	5.10	4.80	4.40	4.10	3.90	3.60	3.40
44	6.60	6.30	5.90	5.60	5.30	5.10	4.80
45	6.00	5.70	5.40	5.20	4.90	4.60	4.40
46	5.40	4.90	4.50	4.10	3.80	3.40	3.10
47	6.10	5.80	5.50	5.20	5.00	4.80	4.60
48	4.00	3.50	3.10	2.70	2.40	2.00	1.60
49	5.00	4.50	4.10	3.70	3.30	3.00	2.80
50	6.40	5.90	5.60	5.20	4.90	4.60	4.40

Table 6.6: Scaled LHD points for ASET-B Input Variables. [Runs 26 to 50 of $N = 50$] and corresponding Egress Times from 35 to 65 seconds in five second intervals.

Run No.	70s	75s	80s	85s	90s	95s	100s
1	3.80	3.50	3.20	2.90	2.60	2.40	2.10
2	2.70	2.40	2.20	1.90	1.70	1.50	1.20
3	2.80	2.70	2.50	2.30	2.10	2.00	1.90
4	2.00	1.80	1.50	1.20	0.80	0.50	0.10
5	2.70	2.50	2.30	2.20	2.00	1.80	1.70
6	5.00	4.70	4.50	4.20	4.00	3.80	3.50
7	1.90	1.50	1.10	0.80	0.40	0.00	0.00
8	2.60	2.30	2.00	1.80	1.50	1.30	1.00
9	2.70	2.40	2.20	1.90	1.70	1.40	1.20
10	4.10	3.90	3.60	3.30	3.10	2.80	2.60
11	3.90	3.70	3.50	3.40	3.20	3.00	2.90
12	4.20	4.00	3.80	3.60	3.40	3.20	2.90
13	4.10	3.70	3.40	3.20	2.90	2.60	2.40
14	4.70	4.60	4.40	4.30	4.10	4.00	3.90
15	4.90	4.70	4.50	4.20	4.00	3.80	3.60
16	3.40	3.10	2.90	2.60	2.40	2.10	1.90
17	4.00	3.80	3.50	3.30	3.10	2.90	2.60
18	4.30	4.00	3.80	3.60	3.30	3.10	2.90
19	4.70	4.50	4.30	4.10	3.90	3.70	3.50
20	3.40	3.20	3.00	2.80	2.60	2.40	2.20
21	4.40	4.10	3.90	3.60	3.40	3.20	3.00
22	5.20	5.10	4.90	4.70	4.60	4.40	4.30
23	2.20	1.80	1.50	1.10	0.70	0.30	0.00
24	5.50	5.20	5.00	4.70	4.50	4.20	4.00
25	4.00	3.70	3.50	3.20	2.90	2.70	2.40

Table 6.7: Scaled LHD points for ASET-B Input Variables. [Runs 1 to 25 of $N = 50$] and corresponding Egress Times from 70 to 100 seconds in five second intervals.

Run No.	$70s$	$75s$	$80s$	$85s$	$90s$	$95s$	$100s$
26	0.70	0.30	0.00	0.00	0.00	0.00	0.00
27	3.80	3.60	3.40	3.20	3.00	2.80	2.70
28	3.50	3.30	3.10	2.90	2.70	2.60	2.40
29	3.90	3.60	3.30	3.10	2.90	2.60	2.40
30	5.40	5.20	4.90	4.70	4.50	4.30	4.10
31	3.40	3.10	2.80	2.60	2.30	2.10	1.80
32	3.00	2.70	2.40	2.00	1.60	1.20	0.80
33	3.30	3.10	2.80	2.60	2.40	2.10	1.90
34	3.70	3.50	3.30	3.10	3.00	2.80	2.60
35	4.10	3.80	3.60	3.40	3.20	3.10	2.90
36	3.40	3.20	3.00	2.80	2.60	2.40	2.20
37	2.50	2.20	1.80	1.40	1.00	0.60	0.20
38	2.90	2.60	2.30	2.10	1.80	1.60	1.30
39	4.80	4.50	4.20	4.00	3.70	3.50	3.30
40	4.50	4.20	4.00	3.80	3.60	3.40	3.20
41	4.70	4.30	4.10	3.80	3.50	3.20	3.00
42	2.30	2.10	1.80	1.60	1.30	1.10	0.80
43	3.20	3.00	2.80	2.70	2.50	2.30	2.20
44	4.60	4.40	4.20	4.00	3.80	3.60	3.40
45	4.20	3.90	3.70	3.50	3.30	3.10	2.80
46	2.80	2.40	2.10	1.70	1.30	0.90	0.40
47	4.40	4.20	4.00	3.80	3.70	3.50	3.40
48	1.20	0.70	0.20	0.00	0.00	0.00	0.00
49	2.50	2.30	2.00	1.80	1.50	1.30	1.00
50	4.10	3.90	3.70	3.40	3.20	3.00	2.80

Table 6.8: Scaled LHD points for ASET-B Input Variables. [Runs 26 to 50 of $N = 50$] and corresponding Egress Times from 70 to 100 seconds in five second intervals.

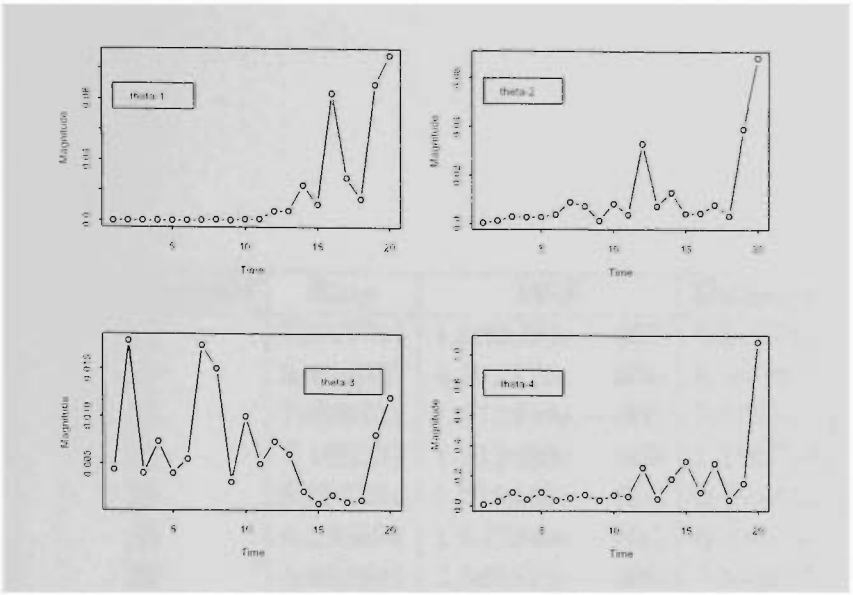

Figure 6.2: Estimated θs for Time in 5 second intervals

5, 10, ..., 100 seconds were generated using the appropriate values of θ and p. In addition the formula from Chapter 2 was used to generate the MSE of prediction. The results are given in columns 2 and 3 of Table 6.9.

To generate a prediction at an intermediate time a weighted cubic spline is used which minimises

$$S(\lambda) = \sum w_i(y_i - f(x_i))^2 + \lambda \int (f''(x))^2 dx$$

where λ is a smoothing parameter varying between 0 and 1 with 1 indicating interpolation. All the predictions are subjected to smoothing where some errors are expected. A smoothed estimate is made using a smoothing cubic spline, and interpolation was not used. Here the weight function w_i, is taken to be proportional to the reciprocal of the MSE.

This was implemented in the S-Plus statistical function smooth.spline. The fourth column of Table 6.9 gives the smoothed estimates of the smoke layer height at times of 5 to 100 seconds in steps of 5 seconds.

Time in second	Blup	MSE	Estimate
5	8.643744	$1.086020e-005$	8.643711
10	8.164935	$4.762126e-005$	8.165607
15	7.699427	$1.073449e-005$	7.699167
20	7.166597	$1.513450e-005$	7.166758
25	6.691224	$1.073449e-005$	6.691403
30	6.299874	$1.035240e-005$	6.299536
35	5.892541	$2.582818e-005$	5.893691
40	5.569429	$9.352271e-006$	5.568876
45	5.179954	$2.466415e-006$	5.180113
50	4.937787	$2.626079e-005$	4.836133
55	4.635693	$1.025419e-005$	1.635979
60	4.392940	$2.594141e-005$	4.392621
65	4.187202	$1.343789e-005$	4.187456
70	4.009393	$2.896289e-005$	4.008583
75	3.745673	$1.790708e-005$	3.745820
80	3.523571	$1.470694e-005$	3.523924
85	3.388127	$1.157005e-005$	3.387755
90	3.182278	$2.876045e-005$	3.182731
95	3.001488	$4.222869e-004$	3.001324
100	2.852575	$6.757982e-004$	2.851340

Table 6.9: Results of analysing Time Trace data using method 1. The BLUP and MSE estimates of the Smoke Layer Height for a heat loss fraction of 0.85, fire height of 2.5 ft, Room Ceiling Height of 9 ft. and Room Floor Area of 150 sq.ft. are based on separate fitting of computer models for each time and the Estimate column is based on using a weighted cubic spline.

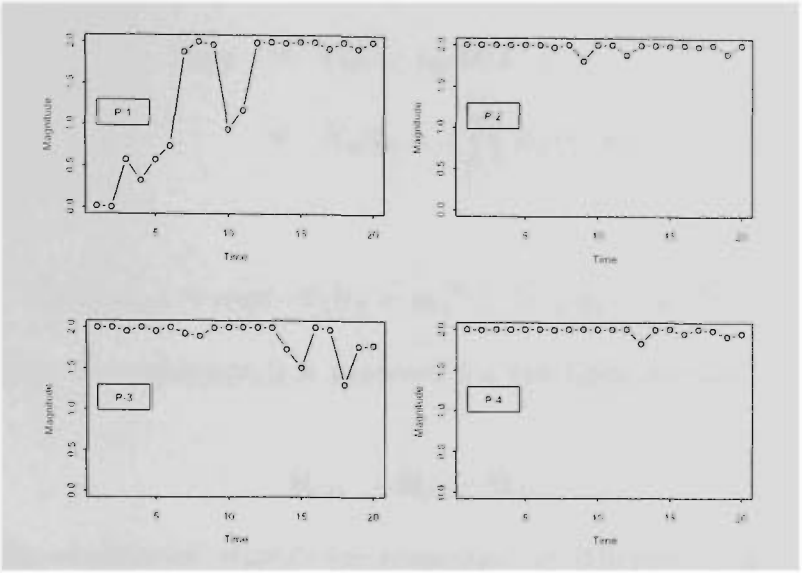

Figure 6.3: Estimated ps for Time in 5 second intervals

6.2.2 Using Time as an Additional Input Factor

When time is used as an additional input factor the observation vector usually will become very large. If at the n design points the output of the computer model is observed at T times, then the observation vector has length nT.

The main difficulty is that in the computation of the likelihood an $nT \times nT$ matrix has to be inverted and the corresponding determinant calculated. However if, as is usually the case, the computer model is observed at the same set of times then the problem can be considerably simplified.

Assume that the computer model is observed at times $T_1, T_2, \ldots, T_n$ and $\mathbf{y}_{T_j}$ denotes the observation vector for time T_j. Then, the combined observation vector is given by

$$\mathrm{vec}(\mathbf{y}_{T_1}, \mathbf{y}_{T_2}, \ldots, \mathbf{y}_{T_n}).$$

The correlation of the zs at time t_0 and input variables $\mathbf{t}$ and the zs at time

u_0 and input variables $\mathbf{u}$ is denoted by

$$\begin{aligned}\mathbf{R}_{DT} &= R_0(t_0, u_0)\mathbf{R}(\mathbf{t}, \mathbf{u}) \\ &= R_0(t_0, u_0)\prod_{k=1}^{d} R_k(t_k, u_k)\end{aligned}$$

where

$$R_k(t_k, u_k) = \exp(-\theta_k|t_k - u_k|^{p_k}) \quad 0 \le p_k \le 2;\ \theta_k \ge 0.$$

In this notation the subscript 0 is reserved for the time variable.

Now

$$\mathbf{R}_{DT} = \mathbf{R}_T \otimes \mathbf{R}_D$$

where $\mathbf{R}_T$, the correlation matrix for responses at different times for a fixed set of conditions, is given by

$$\mathbf{R}_T[i, j] = \exp(-\theta_0|T_i - T_j|^{p_0}) \quad i, j = 1, \ldots, T$$

and $\mathbf{R}_D$, the correlation matrix for the responses at the design points at a fixed time, is given by

$$\mathbf{R}_D[i, j] = \prod_{k=1}^{d} \exp(-\theta_k|x_i - x_j|^{p_k}) \quad i, j = 1, \ldots, n.$$

When calculating the likelihood the following properties of the Kronecker product

$$\mathbf{R}_{DT}^{-1} = \mathbf{R}_T^{-1} \otimes \mathbf{R}_D^{-1}$$

and

$$|\mathbf{R}_{DT}| = |\mathbf{R}_T|\,|\mathbf{R}_D|$$

can be used to simplify the calculations. Using these properties, means that only matrices of size $T \times T$ and $n \times n$ rather than $nT \times nT$ have to be dealt with.

When predicting at $\mathbf{x}_T = (\mathbf{x}, T)^T$ we use

$$\hat{y} = \hat{\beta} + \mathbf{r}_{DT}^T(\mathbf{x})\mathbf{R}_{DT}^{-1}(\mathbf{y} - \mathbf{1}\hat{\beta})$$

where

$$\begin{aligned} \mathbf{r}_{DT}^{T}(\mathbf{x}) &= \mathbf{r}_T(\mathbf{x}) \otimes \mathbf{r}_D(\mathbf{x}) \\ \mathbf{r}_T(\mathbf{x}) &= [R_0(T, T_1), R_0(T, T_2), \ldots, R_0(T, T_n)] \\ \text{and } \mathbf{r}_D &= [\mathbf{R}(\mathbf{x}, \mathbf{x}_1), \mathbf{R}(\mathbf{x}, \mathbf{x}_2) \ldots, \mathbf{R}(\mathbf{x}, \mathbf{x}_n)] . \end{aligned}$$

For the same set of conditions as before, that is a Heat Loss Fraction of 0.85, Fire Height of 2.5 ft, Room Ceiling Height of 9 ft and Room Floor Area of 150 sq. ft, the BLUP predictions of the Smoke Layer Height at 5,10,....100 seconds were generated. Table 6.10 and Figure 6.4 shows a comparison of the predictions against the actual values generated for the ASET-B program. The agreement between these two sets of values is excellent.

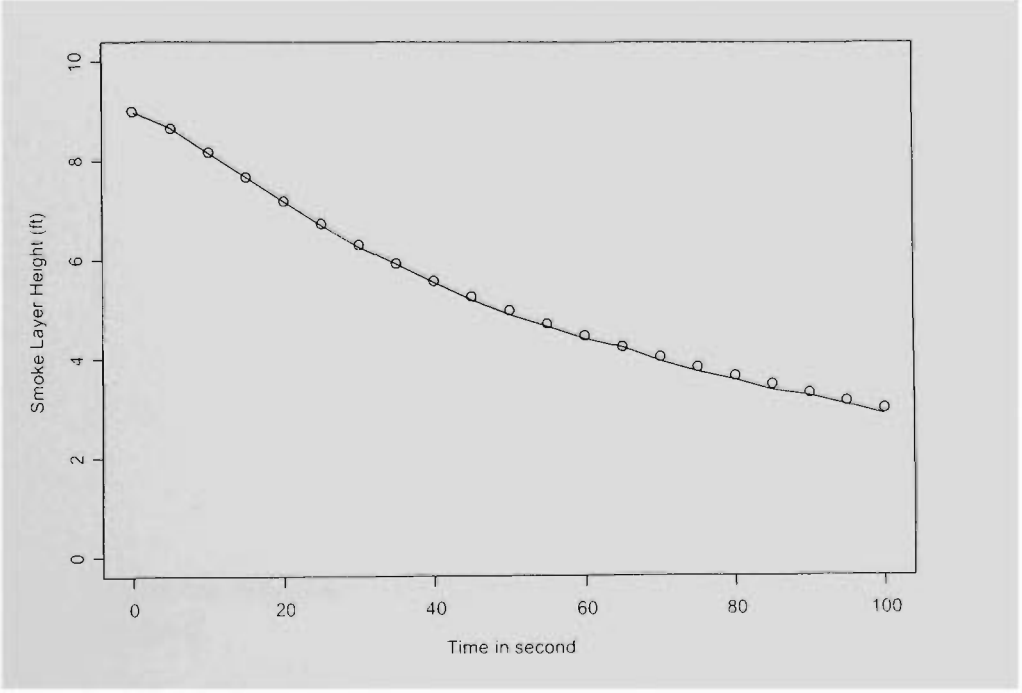

Figure 6.4: Predicted Response versus Actual response at 5 second time intervals using method 2.

Time (s)	Actual	Predicted
0	9.00	8.9822
5	8.65	8.6573
10	8.17	8.1520
15	7.67	7.6683
20	7.18	7.1782
25	6.73	6.7045
30	6.31	6.2632
35	5.92	5.9105
40	5.57	5.5409
45	5.25	5.1862
50	4.97	4.8869
55	4.70	4.6490
60	4.46	4.4030
65	4.23	4.2256
70	4.03	3.9398
75	3.83	3.7335
80	3.65	3.5657
85	3.48	3.3510
90	3.31	3.2570
95	3.15	3.0699
100	2.99	2.8689

Table 6.10: Predicted Response versus Actual response at five second time intervals using method 2.

The MSE of prediction is given by Equation 2.3 and can be simplified as

$$\begin{aligned}
&\mathrm{MSE}(\hat{y}(\mathbf{x}_T)) \\
&= \sigma^2 \left[1 - \begin{pmatrix} 1 & \mathbf{r}_{DT}^T \end{pmatrix} \begin{pmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{R}_{DT} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mathbf{r}_{DT} \end{pmatrix} \right] \\
&= \sigma^2 \left[1 - \begin{pmatrix} 1 & \mathbf{r}_T^T \otimes \mathbf{r}_D^T \end{pmatrix} \begin{pmatrix} 0 & \mathbf{1}_T^T \otimes \mathbf{1}_n^T \\ \mathbf{1}_T \otimes \mathbf{1}_n & \mathbf{R}_T \otimes \mathbf{R}_D \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mathbf{r}_T \otimes \mathbf{r}_D \end{pmatrix} \right].
\end{aligned}$$

Using properties of partitioned matrices

$$\begin{aligned}
&\begin{pmatrix} 0 & \mathbf{1}_T^T \otimes \mathbf{1}_n^T \\ \mathbf{1}_T \otimes \mathbf{1}_n & \mathbf{R}_T \otimes \mathbf{R}_D \end{pmatrix}^{-1} \\
&= \frac{1}{d} \begin{pmatrix} 1 & -\left(\mathbf{R}_T^{-1}\mathbf{1}_T\right)^T \otimes \left(\mathbf{R}_D^{-1}\mathbf{1}_n\right)^T \\ -\left(\mathbf{R}_T^{-1}\mathbf{1}_T\right) \otimes \left(\mathbf{R}_D^{-1}\mathbf{1}_n\right) & \left(\mathbf{R}_T^{-1} \otimes \mathbf{R}_D^{-1}\right) \\ & + \left(\mathbf{R}_T^{-1}\mathbf{J}_T\mathbf{R}_T\right) \otimes \left(\mathbf{R}_D^{-1}\mathbf{J}_n\mathbf{R}_D\right) \end{pmatrix}
\end{aligned}$$

where

$$d = \left(\mathbf{1}_T^T\mathbf{R}_T^{-1}\mathbf{1}_T\right)\left(\mathbf{1}_n^T\mathbf{R}_D^{-1}\mathbf{1}_n\right)$$

and $\mathbf{J}_T$ and $\mathbf{J}_n$ and the $T \times T$ and $n \times n$ matrices respectively with all elements equal to 1. Hence

$$\mathrm{MSE}(\hat{y}(\mathbf{x}_T)) = \sigma^2 \left[1 - \begin{pmatrix} e & \mathbf{f} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{r}_T \otimes \mathbf{r}_D \end{pmatrix} \right]$$

where

$$\begin{aligned}
e &= -\left(\mathbf{r}_T^T \otimes \mathbf{r}_D^T\right)\left(\left(\mathbf{R}_T^{-1}\mathbf{1}_T\right) \otimes \left(\mathbf{R}_D^{-1}\mathbf{1}_n\right)\right)/d \\
&= -\left(\mathbf{r}_T^T\mathbf{R}_T^{-1}\mathbf{1}_T\right)\left(\mathbf{r}_D^T\mathbf{R}_D^{-1}\mathbf{1}_n\right)/d
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{f} &= \Big\{ -\left(\mathbf{R}_T^{-1}\mathbf{1}_T\right)^T \otimes \left(\mathbf{R}_D^{-1}\mathbf{1}_n\right)^T \\
&\quad + \left(\mathbf{r}_T^T \otimes \mathbf{r}_D^T\right)\left(\left(\mathbf{R}_T^{-1} \otimes \mathbf{R}_D^{-1}\right) + \left(\mathbf{R}_T^{-1}\mathbf{J}_T\mathbf{R}_T\right) \otimes \left(\mathbf{R}_D^{-1}\mathbf{J}_n\mathbf{R}_D\right)\right)\Big\}/d \\
&= \Big\{ -\left(\mathbf{R}_T^{-1}\mathbf{1}_T\right)^T \otimes \left(\mathbf{R}_D^{-1}\mathbf{1}_n\right)^T \\
&\quad + \left(\mathbf{r}_T^T\mathbf{R}_T^{-1}\right) \otimes \left(\mathbf{r}_D^T\mathbf{R}_D^{-1}\right) + \left(\mathbf{r}_T^T\mathbf{R}_T^{-1}\mathbf{J}_T\mathbf{R}_T\right) \otimes \left(\mathbf{r}_D^T\mathbf{R}_D^{-1}\mathbf{J}_n\mathbf{R}_D\right)\Big\}/d.
\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{MSE}(\hat{y}(\mathbf{x}_T)) &= \sigma^2 \left\{ 1 - \frac{1}{\left[\left(\mathbf{1}_T^T \mathbf{R}_T^{-1} \mathbf{1}_T\right)\left(\mathbf{1}_n^T \mathbf{R}_D^{-1} \mathbf{1}_n\right)\right]} \left[\left(\mathbf{r}_T^T \mathbf{R}_T^{-1} \mathbf{1}_T\right)\left(\mathbf{r}_D^T \mathbf{R}_D^{-1} \mathbf{1}_n\right)\right.\right. \\ &\quad + \left(\mathbf{1}_T^T \mathbf{R}_T^{-1} \mathbf{r}_T\right)\left(\mathbf{1}_n \mathbf{R}_D^{-1} \mathbf{r}_D\right) - \left(\mathbf{r}_T^T \mathbf{R}_T^{-1} \mathbf{r}_T\right)\left(\mathbf{r}_D^T \mathbf{R}_D^{-1} \mathbf{r}_D\right) \\ &\quad \left.\left. - \left(\mathbf{r}_R^T \mathbf{R}_T^{-1} \mathbf{J}_T \mathbf{R}_T \mathbf{r}_T\right)\left(\mathbf{r}_D^T \mathbf{R}_D^{-1} \mathbf{J}_n \mathbf{R}_D^{-1} \mathbf{r}_D\right)\right] \right\}.\end{aligned}$$

6.3 Conclusion

Many computer experiments give as a response a trace over time. Two different methods for analysing such time traces have been presented.

In the first method each time set is analysed separately, and an individual model is estimated for each time point. To obtain predictions for untried inputs, individual predictions at the experimental times are made using the models and then a smoothed estimate is made using a smoothing cubic spline. When fitting the spline the reciprocal of the MSEs of the predictions are used as weights. In the example this method appears to work well.

An alternative method uses time as an additional input factor. The apparent disadvantage of the much increased dimension of the observation vector can, at least for the product correlation family considered here, be overcome by using Kronecker products in the calculation of the likelihood. One advantage of the second method is that the usual formula for the MSE of prediction can be used.

Note that the second method assumes that the outputs at each time can be modelled as realisations of stochastic processes with the same variance-covariance matrix, while with the first method the variance-covariance matrices are free to vary. If this homogeneity assumption is appropriate then the second method gives a unified method which is easily implemented.

Chapter 7

Discussion

7.1 Introduction

In previous chapters a number of investigations have been done on Computer Experiments. These investigations have largely been devoted to practical questions such as how to best estimate parameters, how to summarise results, what sample size should be used, what augmenting runs need to be added and how can time trace responses be analysed easily.

In Chapter 2 the machinery of Computer Experiments was used to develop a surrogate model for the ASET-B computer model. It was shown how a good model giving very accurate results could be generated from 50 runs of the ASET-B program. Explicit expressions for the mean, main effects and two-factor interactions effects were used to summarise the results which showed that the non-additivity in this case was quite small. These results were conformed using Functional Analysis of Variance.

In Chapter 3 four different computer models were studied. For each model, Latin Hypercube of $10, 20, \ldots, 100$ runs were produced and the four computer models were run at inputs corresponding to the design. For each sample size the computer models were modelled as realisations of stochastic processes and the ability to predict at untried inputs was assessed using ERMSE. Analyses of

the data showed that the computer models gave an approximate linear relationship between $ERMSE^{-0.25}$ and n. However, it was found that the slope of this relationship varied considerably from model to model due to limited usefulness.

In Chapter 4 an expression for the prediction covariance was found, generating the well known expression for the prediction variance on MSE. For adding one point at a time choosing the point from a candidate set with the maximum prediction variance appears to work well for the ASET-B model. For adding more than one point at a time the points with the maximum determinant of the prediction variance-covariance matrix were selected.

In Chapter 5 Computer Experiments were studied. For a four dimensional model realisations of a stochastic process with different values of θ and p chosen using a Box-Behken response surface design were generated and the ln(EMSE) was modelled using a quadratic response surface. The relationship shows how the prediction variance depends on θ, p and n on average and the range of values that can be expected. Main effects and joint effect diagrams show the effect of the parameters and the interaction between them. The results are shown to match those of Chapter 3 reasonably well.

In Chapter 6 two methods of analysing Computer Experiments which give a trace of responses over time were studied and contrasted. One method involved fitting a separate model for each time studied. At an untried input, predictions for each time point are produced using the separate models. For predicting at an intermediate time, the predictions are combined using a weighted smoothing spline, with the weights inversely proportional to the MSEs of the predictions for each time point.

The other method involves adding time as an additional input factor. This approach usually gives a much larger dimensional observation vector leading to problems in computing the likelihood. This can be overcome by using the properties of Kronecker products, reducing the problem to much more manageable proportions. However, using this method relies on a stronger assumption than the first method. If this assumption is appropriate then straightforward predictions and estimates of the size of the prediction errors can be computed.

7.2 Limitations

7.2.1 Use of Latin Hypercube

While LHDs are available for all sample sizes there are other designs that may give superior results, certainly for more complex models. Some of the results found for sample sizes will probably give lower bounds on the performance of these other designs.

7.2.2 Relationship between Sample Sizes and EMSE

Although an approximate linear relationship has been established between $\text{ERMSE}^{-0.25}$ and n this relationship is not optimal for all computer models. However, as a compromise it appears to work reasonably well at least for the four computer models considered. Obviously, this compromise may not apply for all computer models. Even if it did, one remark that limits its applicability is that the slope is different depending on the computer model.

7.2.3 Use of only Gaussian Product Correlation Structure

In this thesis only the Gaussian product correlation structure has been used. Other correlation structures may give quite different results. The Gaussian correlation structure is infinitely mean square differentiable for $p = 2$ and not mean square differentiable for $p \neq 2$. Correlation structures that give intermediate levels of mean square differentiability may require different sample sizes.

7.3 Future Work

A number of extensions of the work conducted in this thesis suggest themselves:

7.3.1 Application to other Designs

The work on sample sizes given in Chapter 3 has been limited to Latin Hypercube designs. Similar work should be conducted using other types of designs such as.

random orthogonal arrays, IMSE optimal Latin hypercubes, maximin Latin hypercubes, orthogonal-array based Latin hypercubes, uniform designs, orthogonal Latin hypercubes, Hammersley sequence designs and other designs developed for Computer Experiments. It would be very interesting to see how the relationship between ERMSE and n depends on the types of design. The same applies to the work in Chapter 5.

On the other hand the augmenting methods examined and the methods developed for the analysis of time traces are applicable irrespective of the design used.

7.3.2 Application to other Computer Models

The work on sample sizes was limited to four computer models. Further work of this type would be useful for other computer models of varying dimensions and complexity. Even for large scale computer models the amount of computing could be reduced. For example, if a surrogate model based on 50 runs of the actual code was developed, then the surrogate model could be used to see what would have been the effect of a different sample size on the average prediction error as measured by the ERMSE. These simulation studies would assist experimenters in designing further experiments on the same code but also future experiments on different code.

7.3.3 Prior Estimates of θ and p

Fitting the model by maximum likelihood is made much easier by good starting values of the θ and p parameters. In the absence of good starting values, random starts can be used. Methods of determining good starting values would be very useful.

Prior to collecting the data it would also be useful to have an idea of likely ranges for the θ and p values, in order to assist the determination of an appropriate sample size. Methods for doing this would be of great benefit to experimenters.

7.3.4 Model Diagnostics for Time Trace Data

Two methods of analysing time trace data from computer experiments were given in Chapter 6. The method based on the use of Kronecker products assumes much more about the covariance matrix than does the other method. Diagnostics to see whether this is a reasonable assumption need to be developed. More generally, diagnostics tests and graphs for the correlation structure would be useful.

7.3.5 Multivariate Responses

Most Computer Experiments give as outputs more than one response while the results in Chapter 6 show how time trace data can be conveniently analysed. It would be useful to examine the benefits of analysing more than one response at a time.

Bibliography

ARCHER, G.E.B., SALTELLI, A., and SOBOL, I.M. (1997). "Sensitivity measures, ANOVA-like techniques and the use of bootstrap". *Journal of Statistical Computation and Simulation*, *58*, 99–120.

BARTELL, S.M., GARDNER, R.H., O'NEILL, R.V., and GIDDINGS, J.M. (1983). "Error analysis of predicted fate of anthracene in a simulated pond". *Environmental Toxicology and Chemistry*, *2*, 19–28.

BARTELL, S.M., LANDRUM, J.P., GIESY, J.P., and LEVERSEE, G.J. (1981). "Simulated transport of polycyclic aromatic hydrocarbons in artificial streams". In W.J. Mitch, R.W. Bosserman, and J.M. Klopatek, editors, *Energy and Ecological Modelling*, 133–143. Elsevier, New York.

BATES, R.A., BUCK, R.J., RICCOMAGNO, E. and WYNN, H.P. (1996). "Experimental Design and Observation for Large Systems". *Journal of the Royal Statistical Society, Series B*, *58*, 77–94.

BIRK, D.M. (1991). *An Introduction to Mathematical Fire Modeling*. Technomic Publishing: Lancaster, PA.

BOOKER, A. J. (1996). "Case Studies in Design and Analysis of Computer Experiments", *ASA Proceedings of the section on Physical and Engineering Sciences*, 244–248.

BOOKER, A. J., KARIMI, K. J., and MONG, A. (1996). "Modeling, simulation and verification of large DC power electronic systems". In *Proceedings of the* 27^{th} *IEEE Power Electronics Specialists Conference*, 1731–1738.

BOX, G.E.P., and BEHNKEN, D.W. (1960). "Some new three level designs for the study of quantitative variables". *Technometrics*, **2**, 455–475.

BOX, G.E.P., and DRAPER, N.R. (1987). *Empirical Model-Building and Response Surfaces*, John Wiley and Sons.

BOX, G.E.P., and FUNG, C.A. (1986). "Studies in Quality Improvement: Minimising Transmitted Variation by Parameter Design". In *Center for Quality and Productivity Improvement, University of Wisconsin-Madison, Report No. 8*, 1–37.

BOX, G.E.P., W.G. HUNTER, and J.S. HUNTER. (1978). *Statistics for Experimenters*, John Wiley and Sons., New York.

BOX, M.J., and DRAPER, N.R. (1971). "Factorial Designs, the $|X'X|$ criterion, and some related matters", *Technometrics, 13*, 731–742.

BUTLER, N.A. (2001). "Optimal and orthogonal Latin hypercube design for computer experiments", *Biometrika, 88*, 847–857.

CARLING, K. (1995). "Starting values for the iterative maximum likelihood estimator in survival analysis", *Journal of Applied Statistics, 22*, 531–535.

CHANG, P.B., WILLIAMS, B.J., SANTER, T.J., NOTZ, W.I., and BARTEL, D.L. (1999). "Robust Optimization of Total Joint Replacements Incorporating Environmental Variables", *Journal of Biomechanical Engineering, 121*, 304–310.

COOPER, L.Y., and FORNEY, G. P.(1990). "The Consolidated Compartment Fire Model (CCFM) Computer Code Application CCFM. Vents-Part I: Physical Basis", *National Institute of Standards and Technology, NISTIR 4342*, Gaithersburg, MD.

COOPER, L., (1980). " Estimating safe available egress time from fires," Technical report 80-2172, Nat. Bur. Stand. (U.S.).

CRESSIE, N. (1986). "Kriging Non stationary Data", *Journal of the American Statistical Association, 81*, 625–634.

CRESSIE, N. (1993). "Statistics for spatial data", Wiley, New York.

CRAIG, P.S., GOLDSTEIN, M., SEHEULT, A.H. and SMITH, J.A. (1996). "Bayes linear strategies for matching hydrocarbon reservoir history". In Bayesian Statistics 5, Ed. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, pp. 65–69. Oxford University Press.

CURRIN, C., MITCHELL, T., MORRIS, M., and YLVISAKER, D. (1991). "Bayesian Prediction of Deterministic Functions, with Applications to the Design

and Analysis of Computer Experiments". *Journal of the American Statistical Association, 86*, 953–963.

CURRIN, C., MITCHELL, T., MORRIS, M., and YLVISAKER, D. (1988). "A Bayesian Approach to the Design and Analysis of Computer Experiments". *Technical Report 6498*, Oak Ridge National Laboratory.

DYKSTRA, O. (1971). "The Augmentation of Experimental Data to Maximize $|X'X|$". *Technometrics, 13*, 682–688.

EASTERLING, R.G. (1989b). "**Comment:** Design and Analysis of Computer Experiments", *Journal of Statistical Sciences, 4*, 425–427.

EVANS, D.D., and STROUP, D. W. (1985). "Methods to Calculate the Response Time of Heat and Smoke Detectors Installed Below Large Unobstructed Ceilings", *Fire Technology, 22*, 54–65.

FANG, K.T. and WANG, Y. (1994). *Number-theoretic Methods in Statistics*, Chapman and Hall, New York.

FISHER, R.A. (1935). *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.

GOLDBERGER, A. S. (1962). "Best Linear Unbiased Prediction in the Generalized Linear Regression Model", *Journal of the American Statistical Association, 57*, 369–375.

HENDERSON, C. R. (1975b). "Use of all relatives in intraherd prediction of breeding values and producing abilities", *Journal of Dairy Sciences, 58*, 1910–1916.

HO, C.P., HANSEN, S.E., and FAHEY, P.M. (1984). "Suprem III - A program fro integrated circuit process modeling and simulation", *TR-SEL84-1*, Stanford Electronics Laboratories.

HUANG, J.Z., KOOPERBERG, C., STONE, C.J., and TRUONG, Y.K. (2000). "Functional ANOVA Modeling for Proportional Hazards Regression", *The Annals of Statistics, 28*, 961–999.

JOHNSON, M.E., MOORE, L.M., and YLVISAKER, D. (1990). "Minimax and Maxmin Distance Designs", *Journal of Statistical Planning and Inference, 26*, 131–148.

JONES, D., SCHONLAU, M., WELCH, W., (1998). "Efficient Global Optimization of Expensive Black-Box Functions", *Journal of Global Optimization, 13*, 455–492.

IMAN, R.L., and CONOVER, W.J. (1982). "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables", *Communication in Statistics, part B, 11*, 311–334.

IMAN, R.L. and CONOVER, W.J. (1980). "Small Sample Sensitivity Analysis Techniques for Computer Models, with an Application to Risk Assessment", *Communication in Statistics, -Theory and Methods, A9(17)*",1749–1842.

KALAGNANAM, J. R. and DIWEKAR, U. M. (1997). "An Efficient Sampling Technique for Off-Line Quality Control", *Technometrics, 39*, 308–319.

KENNEDY, M.C. and O'HAGAN, O. (2000). "Predicting the output from a complex computer code when fast approximations are available", *Biometrika, 87*, 1–13.

KLEIJNEN, J.P.C. and SARGENT, R.G. (2000). "A methodology for fitting and validating metamodels in simulation", *European Journal of Operational Research, 120*, 14–29.

KLEIJNEN, J.P.C.(2000). "Strategic Drirections in Verification, Validation, and Accreditation Research": a personal view, *Proceeding of the 2000 Winter Simulation Conference* (edited by J.A. Jones, R.R. Barton, K. Kang and P.A. Fishwick), pp. 909–916.

KLEIJNEN, J.P.C. and van BEERS, W.C.M. (2003). "Application-driven Sequential Design for Simulation Experiments: Kriging Metamodeling", *Journal of the Operational Research Society*, ***under review***.

KOEHLER, J.R. and OWEN, A.B. (1996). "Computer Experiment", Handbook of Statistics, Volume 13: edited by S. Ghosh and C.R. Rao, Elsevier, Amsterdam, pp. 261–308.

LAW, A.M. and KELTON, W.D. (2000). "Simulation modeling and Analysis", Third Edition, McGraw-Hill, Boston.

LOH, W.L. (1996). "A Combinational Central Limit Theorem for Randomized Orthogonal Array Sampling Designs", *The Annals of Statistics, 24*,1209–1224.

LUCAS, J. M. (1996). "Letters to the Editor. Comments on Computer Experiments". *Technometrics, 38*. 197 199.

MATHERON, G. (1963). "Principles of geostatistics". *Economic Geology, 58*. 1246 1266.

McKAY, M.D., CONOVER, W.J., and BECKMAN, R.J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code",*Technometrics, 21*. 239 245.

MECKESHEIMER, M., BARTON, R.R., SIMPSON, T.W. and BOOKER, A.J. (2002). "Computationally inexpensive metamodel assessment strategies". *AIAA Journal, 40*. 2053 2060.

MERTENS, B.J.A. (2001). "Downdating: interdisciplinary research between statistics and computing". *Statistica Neerlandica, 55*. 358 366.

MILLER, D., and FRENKLACH, M. (1983). "Sensitivity analysis and parameter estimation in dynamic modeling of chemical kinetics". *International Journal of Chemical Kinetics, 15*. 677 696.

MORRIS, M.D., MITCHELL, T.J. (1995). "Exploratory Designs for Computational Experiments". *Journal of Statistical Planning and Inference, 43*. 381 402.

NASSIF, S.R., STROJWAS, A.J., and DIRECTOR, S.W. (1984). "FABRICS II: a statistically based IC' fabrication process simulator". *IEEE Trans. Computer-Aided Design, 3*. 40 46.

OLSSON, A., SANDBERG, G., and DAHLBLOM, O. (2003). "On Latin hypercube sampling for structural reliability analysis". *Structural Safety, 25*. 47 68.

OWEN, A. B. (1994a). "Controlling Correlations in Latin Hypercube Samples". *Journal of American Statistical Association, 89*. 1517 1522.

OWEN, A. B. (1994b). "Lattice Sampling Revisited: Monte Carlo Variance of Means Over Randomized Orthogonal Arrays". *The Annals of Statistics, 22*. 930 945.

OWEN, A. B. (1992a). "A Central Limit Theorem for Latin Hypercube Sampling". *Journal of the Royal Statistical Society, B 54*. 541 555.

OWEN, A.B. (1992b). "Orthogonal Arrays for Computer Experiments, Integration and Visualization". *Statistica Sinica, 2*, 439 452.

PARK. J.S. (1994)."Optimal Latin-Hypercube Designs for Computer Experiments". Parkinson. A. and al.. c.. 1998."OptdesX: A Software System for Optimal Engineering Design. "User's Manual. Release 2.0.4. Design Synthesis. Inc.. Provo. UT.

PALMER. K. and TSUI. K. (2001). "A minimum bais Latin hypercube design". *IIE Transaction, 33.* 793 808.

PATTERSON. H.D. (1954). "The Error of Lattice Sampling". *Journal Royal Statistical Society. Series B 16.*140 149.

PRIBADI. Y., VOETEN. J.P.M. and THEELEN. B.D. (2001). "Reducing Markov Chains for Performance Evaluation". *Proceedings of Progress.* 173 179.

RAGHAVRAO. D. (1971). *Construction and Combinational Problems in Design of Experiments.* John Wiley. New York.

GRAY. R.M. (2001). "Probability. Random Processes and Ergodic Properties". Springer Verlag.

ROBINSON. G. K. (1991). "That BLUP Is a Good Thing: The Estimation of Random Effects". *Statistical Sciences. 6.*15 51.

RIPLEY. B.D. (1981). *Spatial Statistics.* John Wiley. New York.

SACKS. J.. SCHILLER. B.S. and WELCH. W.J. (1989a). "Designs for Computer Experiments".*Technometrics. 31.* 41 47.

SACKS. J.. WELCH. W.J. MITCHELL. T.J.. and WYNN. H.P. (1989b). "Design and Analysis of Computer Experiments". *Journal of Statistical Sciences. 4.* 409 435.

SAHAMA. T.R.. and DIAMOND. N.T. (2001). "Sample Size Considerations and Augmentation of Computer Experiments". *J. Statist. Comput. Simul.. 68.* 307 319.

SALTELLI. A., TARANTOLA. S.. and CHAN. K.P.S. (1999). "A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output". *Technometrics. 41.* 39 56.

SARGENT. R.G.. GLASOW. P.A.. KLEIJNEN. J.P.C.. LAW. A.M.. McGREGOR. I. and YOUNGBLOOD. S. (2000). "Strategic Drirections in Verification. Validation. and Accreditation Research": a personal view. *Proceeding of the 2000*

Winter Simulation Conference (edited by J.A. Jones, R.R. Barton, K. Kang and P.A. Fishwick), pp. 909 916.

SCHONLAU, M. (1997). "Computer Experiments and Global Optimization" Ph.D. Dissertation, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, CA.

Schonlau, M., Welch, W., and Jones, D. (1998). "Global Versus Local Search in Constrained Optimization of Computer Models", *In N. Flournoy, W.F. Rosenberger, and W.K. Wong (Eds.), New Developments and Applications in Experimental Design, Institute of Mathematical Statistics, Hayward, California, 34,* 11 25.

SHARIFZADEH, S., KOEHLER, J.R., OWEN, A.B., and SHOTT, J.D. (1989). "Using simulators to model transmitted variability in IC manufacturing", *IEEE Trans. Semicond. Manufact. 2(3),* 82 93.

SHEWRY, M.C., and H.P. WYNN, (1987). "Maximum Entropy Sampling", *Journal of Applied Statistics, 14,* 165 170.

SIMPSON, T.W. (1998). "A Concept Exploration Method for Product Family Design", Ph.D. Dissertation, School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA.

STEIN, M. (1987). "Large Sample Properties of Simulations using Latin Hypercube sampling", *Technometrics, 29,* 143 151.

TAGUCHI, G. (1986). "Introduction to Quality Engineering: Designing Quality Into Products and Processes", White Plains, NY: Kraus International Publications.

SPOSITO, G. (1997). "Ergodicity and the "scale effect" ", *Advances in Water Resources, 20,* 309 316.

STONE, M. (1974). "Cross-validatory choice and assessment of statistical predictions", *Journal Royal Statistical Society, Series B, 36,* 111 147.

TANG, B. (1993). "Orthogonal Array-Based Latin Hypercube", *Journal of the American Statistical Association, 88,* 1392 1397.

VENABLES, W.N., and RIPLEY, B.D. (1995). *Modern Applied Statistics with S-plus*, Springer-Verlag, New York.

WALTON, W.D. (1985). *ASET-B: A Room Fire Program for Personal Computers, NBSIR 85-3144-1*. National Bureau of Standards. Washington. D.C.

WELCH. W.J., BUCK. R.B., SACKS. J., WYNN. H.P., MORRIS. M.D., and SCHONLAU. M. (1996). "Letters to the Editor. Response to James M. Lucas",*Technometrics. 38*. 199 203.

WELCH, W.J., BUCK, R.J., SACKS. J., WYNN. H.P., MITCHELL. T.J. and MORRIS, M.D. (1992). "Screening, Predicting and Computer Experiments". *Technometrics. 34*, 15 25.

WELCH, W.J. and SACKS, J. (1991). "A System for Quality Improvement Via Computer Experiments". *Communication in Statistics. -Theory and Methods. 20*. 477 495.

YE. K.Q. (1998). "Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments". *Journal of the American Statistical Association. 93*. 1430 1439.

YE. K.Q., LI. W. and SUDJIANTO. A. (2000). "Algorithmic construction of optimal symmetric Latin hypercube designs". *Journal of Statistical Planning and Inference. 90*. 145 159.